

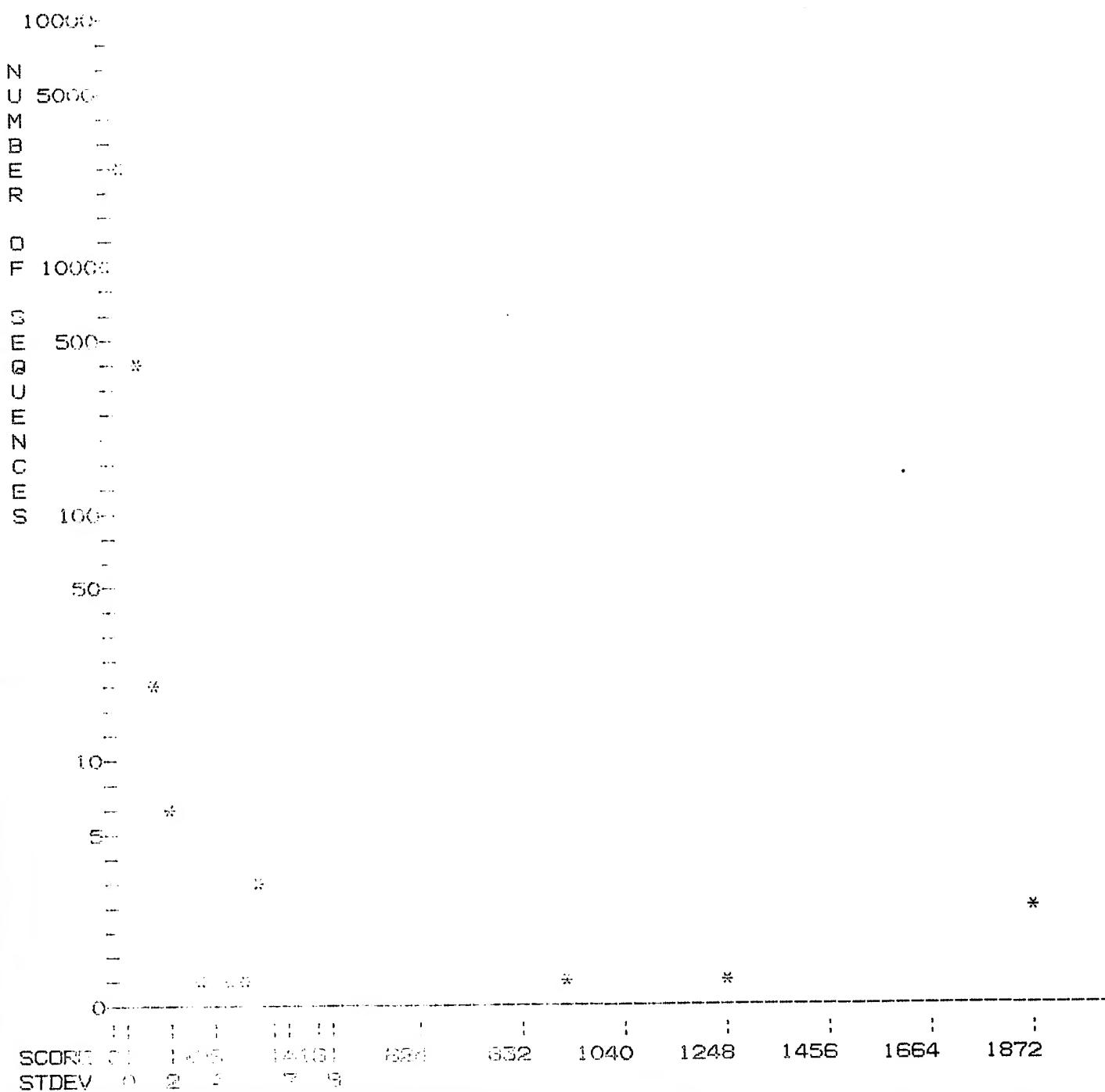
Kunz
07/158652
Claim 32 U

Results file: Kunz-158-Cl32umb1.res made by sheppard on Thu 8 Mar 90 9:51:34-PST.

EMBL

Query sequence being compared: KUNZ-158-CL32. SEQ
Number of sequences searched: 3460
Number of scores above cutoff: 10

Results of the initial comparison of KUNZ-158-CL32. SEQ with:
Data bank: EMBL 21, all entries



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	-1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	147		
Randomization group	0		
Initial scores (0) score	20	Alignments to save	10
Optimized scores (0) score	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	40	38	46.71
Times:	CPU		Total Elapsed
	00:07:21.04		00:21:31.00
Number of references:	5125998		
Number of sequences searched:	3460		
Number of scores above cutoff:	10		

The scores below are sorted by initial score.

Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.					
		Length	Score	Score	Sig.	Frame	
**** 38 standard deviations above mean ****							
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2186	39.22	0	
2. HTVH3CII	Human T-cell lymphotropic virus	9749	1872	2186	39.22	0	
**** 25 standard deviations above mean ****							
3. H1VELING	Human lymphadenopathy virus (E)	9176	1246	1904	25.82	0	
**** 18 standard deviations above mean ****							
4. HEMMADE	Human lymphadenopathy virus (M)	9229	916	2060	18.76	0	
**** 5 standard deviations above mean ****							
5. H1VERDOK	Human immunodeficiency virus t	9671	306	1194	5.70	0	
6. REHTLV4C	HTLV-3 (HTLV-4) partial provir	5391	299	1257	5.55	0	
7. RE1IVAXA	Simian immunodeficiency virus	9264	294	1250	5.44	0	
8. RE1IVMDL	Simian immunodeficiency virus	9646	290	1247	5.35	0	
**** 4 standard deviations above mean ****							
9. RE1IV2P1	Simian immunodeficiency virus	1142	258	565	4.69	0	
**** 3 standard deviations above mean ****							
10. M15127	Figure 1. Structure of the art	306	184	298	3.08	0	

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.					
		Length	Score	Score	Sig.	Frame	
**** 0.00 standard deviations above mean ****							
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2186	0.00	0	
2. HTVH3CII	Human T-cell lymphotropic virus	9749	1872	2186	0.00	0	

FT	SITE	453	453	U3 region
FT	CAP	454	454	cap site
FT	SITE	454	551	R region
FT	SITE	552	634	U5 region
FT	INVREP	633	634	inverted repeat
FT	SITE	635	653	tRNA binding site (tRNA-Lys)
FT	CDS	787	1182	gag p17
FT	CDS	787	2321	gag precursor polypeptide
FT	CDS	1183	2321	gag p24 and gag p15 for major capsid protein and for put. retroviral nucleic acid binding protein (NBP)(ref. 2) (boundaries not defined)
FT				direct repeat
FT	RPT	1968	2002	direct repeat
FT	RPT	2031	2065	pol precursor polypeptides
FT	CDS	2081	5125	put. protease at 5' terminus
FT				reverse transcriptase
FT				put. endonuclease at 3' terminus
FT	RPT	2128	2163	direct repeat
FT	RPT	2164	2176	direct repeat
FT	CDS	5040	5648	SOR short open reading frame
FT				pot. vestigial env gene
FT	CDS	6323	8821	env-lor precursor polypeptide
FT	CDS	6323	8821	envelope glycoprotein
FT	SITE	7782	7787	put. peptide cleavage site
FT	CDS	7787	8821	put. lor transmembrane protein
FT	SITE	9098	9103	poly purine stretch
FT	SITE	9115	9567	U3 region
FT	RPT	9116	9748	long terminal repeat
FT	SITE	9568	9665	R region
FT	SITE	9641	9646	polyadenylation signal
FT	SITE	9668	9743	U5 region
FT	INVREP	9747	9748	inverted repeat

XX

SQ Sequence 9748 BP; 3431 A; 1781 C; 2368 G; 2168 T; 0 other;

Initial Score = 1872 Optimized Score = 2186 Significance = 0.00
 Residue Identity = 85% Matches = 2255 Mismatches = 157
 Gaps = 110 Conservative Substitutions = 0

X	10	20	30	40	50	60
AAGAAT-CAG---	AAGACAGTGGCAATGAGAGTGAAAGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA					
AATAGACAGETTAATTAGATAGACTAATAGAAAAC-ACCGAGAAGACAGTGGCAAT----GAGAGTGAAGGGAGAA						
X 6210	6220	6230	6240	6250	6260	6270

70	80	90	100	110	120		
AATGAGGGCAC-CATGCACCTTGCGTATTTCATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT							
ATATCAGCAUTTATGAGATGGGGTGCGAGATGGGCCACCATGCTCCCTGGGATGTTGATGATCTGTAGTGC							
6270	6280	6290	6300	6310	6320	6330	6340

130	140	150	160	170	180	
CACAA-TCTTATTAATGGTAC---CT---GTGTGGAA-----GGAAGCAA-CCACCA-CTCTATTTGTG						
TACAGAAAATTAATGTCAGATCTATTATGGGTACCTGTGTGGAAAGGAAGCAACCACCACTCTATTTG						
6350	6360	6370	6380	6390	6400	6410

190	200	210	220	230	240
CATGAGATTTTAAAGGTTATGATA ---CAGAGG-TACATA-----AT--GTTTGGGCCACACATGCCTG--T					
-----TCT---GAGAGATCTTAAAGGATATGATACAGAGGTACATAATGTTGGGC-CACA--CATGCCT					
6420	6430	6440	6450	6460	6470

250 260 270 280 290 300 310
 GTCACGAGATGCGGAGGGAAAGTAGTATTGGTAATGTGACAGAAAATTTAACATGTGGAAA-
 TTT
 GTAAATGTCGACA---GAAAT
 3200 3400 3500 3600 3700
 -ATGCATTTTGAACTGATGTTTGTGAAATAATCAG-ITTATG-GGATCAAAGCCTAAAGCCATGTG-T
 TTT
 TTTTAAAGTGTGAACTGATGAAATAATCAGATGAGGGATATAA---TCAGTTTATGGGAT
 6530 6630 6650 6660 6670 6680 6690
 380 390 400 410 420 430 440
 AAAAATTTGGCGCAATCTTGTGTTTTGTTGTTGACTGATTTGG---GGAATGCTACTAAT---ACCAA
 TTT
 CAGGTTGTTGAAATTAACCCCCACTCTGTGTTAGTTAAAGTGC-ACTGATTGAAAGAA
 6600 6610 6620 6630 6640 6650 6660
 450 460 470 480 490 500 510
 TACTGATTTTGAACTGATGAAATAATCAGATGATGAAAGGAGAGATAAAAAACTGCTCTTCAA
 TTT
 TGAATGCTACTAATGAAATAATGAAATGATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTCAA
 6670 6680 6690 6700 6710 6720 6730
 520 530 540 550 560 570 580
 TATGCGGCGAACTGATGAAATGAGTGTGAAAGGAAATATGCACTTTTATAAACTTGATATAATACCAAT
 TTT
 TATGCGGCGAACTGATGAAATGAGTGTGAAAGGAAATATGCACTTTTATAAACTTGATATAATACCAAT
 6740 6750 6760 6770 6780 6790 6800
 590 600 610 620 630 640 650
 AGATTAATGATGAAATGAACTGACATTGCAATTTGTAACACACCTCAGTCATTACACAGGCCCTGTCCAAAGGT
 TTT
 AGATTAATGATGAAATGAACTGACATTGCAATTTGTAACACACCTCAGTCATTACACAGGCCCTGTCCAAAGGT
 6810 6820 6830 6840 6850 6860 6870
 660 670 680 690 700 710 720 730
 ATCCCTTGGCGAAATGCAATGATGAACTGACATTGCAATTTGCAACGGCTGGTTTTGCGATTCTAAATGTAATAATAAGAC
 TTT
 ATCCCTTGGCGAAATGCAATGATGAACTGACATTGCAACGGCTGGTTTTGCGATTCTAAATGTAATAATAAGAC
 6890 6900 6910 6920 6930 6940
 740 750 760 770 780 790 800
 GTTCAATGAAAGGGCAATGATGAAATGCAACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATC
 TTT
 GTTCAATGAAAGGGCAATGATGAAATGCAACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATC
 6950 6960 6970 6980 6990 7000 7010 7020
 810 820 830 840 850 860 870
 AACCTGAACTGAAATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGATCTGCCAATTCACAGACAA
 TTT
 AACCTGAACTGAAATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGATCTGCCAATTCACAGACAA
 7030 7040 7050 7060 7070 7080 7090
 880 890 900 910 920 930 940
 TGCTAAAGGAGATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGATCTGCCAATTCACAGACAA
 TTT
 TGCTAAAGGAGATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGATCTGCCAATTCACAGACAA
 7100 7110 7120 7130 7140 7150 7160
 950 960 970 980 990 1000 1010
 AACCTGAACTGAAATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGGAAATATGAGACA
 TTT
 AACCTGAACTGAAATGATGAACTGACATTGCAAGGAGAGTAGTAATTAGGAAATATGAGACA
 7170 7180 7190 7200 7210 7220 7230

1830 1836 1840 1850 1860 1870 1880
 CCTGAGGATTTCTTGGGTGCTTCAAACCTCAATTGACCACTCTGTGCCTTGGAAATGCTAGTTGGAGTAA
 CCTGAGGATTTCTTGGGTGCTTCAAACCTCAATTGACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAA
 8030 8040 8050 8060 8070 8080 8090 8100
 1890 1900 1910 1920 1930 1940 1950
 TAATGCTGAGGACAACTTGGATTAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACAATTACACAAG
 TAATGCTGAGGACAACTTGGATTAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACAATTACACAAG
 8110 8120 8130 8140 8150 8160 8170
 1960 1970 1980 1990 2000 2010 2020
 CTTCATACATTCTTAACTGAAATCGAAPPACCAGCAAGAAAAGAATGAACAGAAATTATTGGAAATTAGA
 CTTCATACATTCTTAACTGAAATCGAAPPACCAGCAAGAAAAGAATGAACAGAAATTATTGGAAATTAGA
 8180 8190 8200 8210 8220 8230 8240
 2030 2040 2050 2060 2070 2080 2090
 TAAATGGGCAAACTTGAGATTGTTAACATAACAAATTGGCTGTGGTATATAAAAATTTCATAATGAT
 TAAATGGGCAAACTTGAGATTGTTAACATAACAAATTGGCTGTGGTATATAAAAATTTCATAATGAT
 8250 8260 8270 8280 8290 8300 8310
 2100 2110 2120 2130 2140 2150 2160 2170
 AGTTCAGGAGCTTGGTGGTAAATAATGTTTGCTGTACTTCTATAGTAGAGTTAGGCAGGGATA
 AGTTCAGGAGCTTGGTGGTAAATAATGTTTGCTGTACTTCTGTAGTAGAGTTAGGCAGGGATA
 8320 8330 8340 8350 8360 8370 8380
 2180 2190 2200 2210 2220 2230 2240
 TTCAACATTTGCTTAAACACCTCCAAAGGGAGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
 TTCAACATTTGCTTAAACACCTCCAAAGGGAGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
 8390 8400 8410 8420 8430 8440 8450 8460
 2250 2260 2270 2280 2290 2300 2310
 AGGTTCAGGAGCTTGGTGGTAAATAATGTTTGCTGTACTTCTGTAGTAGCTTACAGTATCTGGGACGATCT
 AGGTTCAGGAGCTTGGTGGTAAATAATGTTTGCTGTACTTCTGTAGTAGCTTACAGTATCTGGGACGATCT
 8470 8480 8490 8500 8510 8520 8530
 2320 2330 2340 2350 2360 2370 2380
 AGCAGGCTTGTACCTCTTCAAGTACCAACGGTTGAGAGAGCTTACTCTTGATTGTAACGAGGATTGTGGAAC
 AGCAGGCTTGTACCTCTTCAAGTACCAACGGTTGAGAGAGCTTACTCTTGATTGTAACGAGGATTGTGGAAC
 8540 8550 8560 8570 8580 8590 8600
 2390 2400 2410 2420 2430 2440 2450
 TTCTTGGGAGGAGGGGGTGGGAAACCTCAAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACTAA
 TTCTTGGGAGGAGGGGGTGGGAAACCTCAAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAGCTAA
 8610 8620 8630 8640 8650 8660 8670
 2460
 AG
 AA
 AG
 X

2. KUNZ-158-0139-SEU
 HTLV-III Human T-cell lymphotropic virus type III, complete

ID: HU1H39- standard: RNA: 9749 BP.
 XX

AC KOP010; KUB004; KO2001
XX
DT 08-SEP-1987 (am correction)
DT 08-SEP-1987 (am correction)
DT 08-SEP-1987 (am correction)
DT 08-SEP-1987 (am correction)
DT 01-SEP-1987 (am correction)
DT 23-JUN-1987 (editor modifications)
DT 29-OCT-1988 (incorporated)
XX
DE Human T-cell lymphotropic virus type III, complete reference genome
(isolates HXB2, HXB3, BH10, BHS and BH8 of HTLV-III DNA).
XX
KW acquired immune deficiency syndrome; complete genome; env gene;
KW gag gene; long terminal repeat; pol gene; polyprotein; provirus;
KW reverse transcriptase; TAR protein; trans-activator.
XX
OS Human T-cell lymphotropic virus type III
OC Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN 111 (bases 1-653, 8118-8749)
RA Starcich B. R., Ratner L., Josephs S. F., Okamoto T., Gallo R. C.,
RA Wong-staal F.;
RT "Characterization of long terminal repeat sequences of htlv-iii";
RL Science 237:539-540(1985).
XX
RN 111 (bases 1-6749)
RA Ratner L., Haseltine W., Patarca R., Livak K. J., Starcich B. R.,
RA Josephs S. F., Doren E. R., Rafalski J. A., Whitehorn E. A.,
RA Baumleiter K., Ivanoff L., Petteway S. R. Jr., Pearson M. L.,
RA Leutenberger J. A., Papas T. S., Ghrayeb J., Chang N. T., Gallo R. C.,
RA Wong-staal F.;
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
RL Nature 313:277-284(1985).
XX
RN 111 (bases 508-8668) exons only, tat mRNA
RA Arya S. K., Guo G., Josephs S. F., Wong-staal F.;
RT "Trans-activator genes of human T-lymphotropic virus type III
(HTLV-III)";
RL Science 221:68-73(1985).
XX
RN 111 (bases 5775-6082, 8397-8499)
RA Balirovski J. G., Patarca R., Roser, C. A., Wong-staal F., Haseltine W.;
RT "Location of the trans-activating region on the genome of human
T-cell lymphotropic virus type III";
RL Science 221:74-77(1985).
XX
RN 111 mRNA splice sites
RA Robson A. B., Daugherty D. E., Venkatesan S., Boulukos K. e.,
RA Benn S. I., Folks T. M., Feuring P., Martin M.;
RT "Transcription of novel open reading frames of AIDS retrovirus
during infection of lymphocytes";
RL Science 223:1388-1390(1985).
XX
RN 111 27k antigen cds
RA Allan J. S., Calligan J. E., Lee T. H., McLane M. F., Kanki P. J.,
RA Giacopetti J. E., Essex M.;
RT "A new HTLV-III/LAV encoded antigen detected by antibodies from
AIDS patients";
RL Science 230:810-813(1985).
XX
RN 111 (bases 5772-8833) in hxb-3
RA Chiu J., Ganguly K., Gordon M., Conway R., Schaber M., Kramer R.,
RA Shaw E., Wong-staal F., Reddy R. P.;
RT "HTLV-III env gene products synthesized in E. coli are recognized
by anti-HIV-1 serum in the sera of AIDS patients";
RL Science 230:814-817(1985).

- RL Cell 41: 879-883(1985).
XX
RN LTR promoter and gp120 coding sequences
RA Allan J. S., Colligan J. F., Barin F., McLane M. F., Sodroski J. G.,
RA Rosen C. A., Haseltine W. A., Lee T. H., Essex M.;
RT "Major glycoprotein antigens that induce antibodies in AIDS
RT patients are encoded by HTLV-III";
RL Science 231: 1081-1084(1985).
XX
RN LTR regulatory sequences in the ltr
RA Rosen C. A., Sodroski J. G., Haseltine W. A.;
RT "The location of cis-acting regulatory sequences in the human T
RT cell lymphotropic virus type III (HTLV-III/LAV) long terminal
RT repeat";
RL Cell 41: 813-823(1985).
XX
RN t101 (bases 1-9749)
RA Van Beveren C., Van Beveren C., Van Beveren C.;
RT "Appendix B: HTLV-3/LAV genome";
RL (Im) Weiss R., Teich N., Vermus and Coffin J. M. (eds.);
RNA Tumor Viruses Second Edition: 1102-1148
RL Cold Spring Harbor Laboratory, New York (1985)
XX
RN LTR trans-activator function and tar sequence
RA Rosen C. A., Sodroski J. G., Goh W. C., Dayton A. I., Lippke J.,
RA Haseltine W. A.;
RT "Post-transcriptional regulation accounts for the trans-activation
RT of the human T-lymphotropic virus type III";
RL Nature 318: 555-558(1985).
XX
RN LTR pol coding sequence
RA Manca Veronesi F., Copeland T. D., DeVico A. L., Rahman R.,
RA Orms頓en S., Gallo R. C., Sarnighadharan M. G.;
RT "Characterization of highly immunogenic p66/p51 as the reverse
RT transcriptase of HTLV-III/LAV";
RL Science 231: 1289-1291(1986).
XX
RN LTR the gag sor gene product
RA Ken N. C., Franchini G., Wong-staal F., DuBois G. C., Robey W. G.,
RA Lautenberger J. A., Papas T. S.;
RT "Identification of HTLV-III/LAV sor gene product and detection of
RT antibodies in human sera";
RL Science 231: 1563-1565(1986).
XX
RN LTR pol NH2-terminal region
RA Krieger R. A., Schaberg M. D., Skalka A. M., Ganguly K., Wong-staal F.,
RA Reddy E. P.;
RT "HTLV-III gag protein is processed in yeast cells by the virus
RT pol-proteinase";
RL Science 231: 1580-1584(1986).
XX
RN LTR sor 23k protein
RA Lee T. H., Colligan J. F., Allan J. S., McLane M. F., Groopman J. E.,
RA Essex M.;
RT "A new HTLV-III/LAV protein encoded by a gene found in cytopathic
RT retroviruses";
RL Science 231: 1546-1549(1986).
XX
RN LTR sor 23k protein
RA Sodroski J. G., Goh W. C., Rosen C. A., Tartar A., Portetelle D.,
RA Burny A., Haseltine W.;
RT "Repliactive and cytopathic potential of HTLV-III/LAV with sor
RT gene deletions";
RL Science 231: 1548-1553(1986).
XX
RN LTR sor 1 binding sites in the unspliced region

RA Jones K. A., Kandriana J. T., Luciw P. A., Tjian R.;
RT "Activation of the AIDS retrovirus promoter by the cellular
RT transcription factor, Sp1";
RL *Science* 232:756-759(1986).

XX

RN DNT acceptor and donor splice sites for tat and 27k

RA Arya S. K., Basilio R. C.;

RT "Three novel genes of human T-lymphotropic virus type III: Immune
RT reactivity of their products with sera from acquired immune
RT deficiency syndrome patients";
RL *Proc. Natl. Acad. Sci. U.S.A.* 83:2209-2213(1986).

XX

RN DNT deletion mutants in the tat gene

RA Dayton A. L., Sidranski D. G., Rosen C. A., Goh W. C., Haseltine W. A.;

RT "The trans-activator gene of the human T cell lymphotropic virus
RT type III is required for replication";
RL *Cell* 44:831-847(1986).

XX

RN [20] hypervariable and conserved regions in the env gene

RA Willley R. W., Ruthledge R. A., Dias S., Folks T., Theodore T. S.,
RA Buckler C. E., Martin M. A.;

RT "Identification of conserved and divergent domains within the
RT envelope gene of the acquired immunodeficiency syndrome
RT retrovirus";
RL *Proc. Natl. Acad. Sci. U.S.A.* 83:5038-5042(1986).

XX

RN [21] env cds boundaries

RA Sidranski D. G., Goh W. C., Rosen C. A., Dayton A., Terwilliger E.,
Haseltine W.;

RT "A second post-transcriptional trans-activator gene required for
RT HTLV-III replication";
RL *Nature* 321:412-417(1986).

XX

CC Sequence for [7] was kindly supplied in computer readable form by
CC R. Crawl, 08/17/85. R. Patarca provided sites information and a
CC clean copy for [4], 08/16/85. Acquired immune deficiency syndrome
CC (AIDS) is caused by a retrovirus known by several names, perhaps
CC representing two separate strains: human T-cell lymphotropic
CC virus-III (HTLV-III), whose sequence is given below, and
CC lymphadenopathy-associated virus (LAV) are thought to be one strain
CC differing from AIDS-associated retrovirus type 2 (ARV-2) when
CC overall homology is the criterion. Some reading frame similarities
CC suggest that ARV-2 and LAV are more closely related. All three
CC viruses, whose sequences do not differ by more than 6%, are
CC believed to belong to the C type subfamily Lentiviridae, the "slow"
CC retroviruses. The BH10 sequence differs from BH8 and BH5 by 0.9% in
CC the coding regions and 1.8% in the noncoding regions, and the
CC authors of [21] believe that these are stable variants. The 5' and
CC 3' LTRs of BH10 and BH5 were not fully sequenced; the missing bases
CC (483-675 and 8808-9749) were filled in by [21] from the proviral
CC clone HXB2 [1]. The sequence below is that of BH10 with exception
CC of the variation at position 9197 which allows annotation of the
CC 27K coding sequence. The BH3 sequence spans bases 6033 to 9607, the
CC BH5 sequence spans bases 675 to 6038, and the HXB3 sequence [7]
CC spans bases 5778 to 9843. While this entry is offered as the
CC reference locus for the AIDS retroviral sequence loci, no claim is
CC being made that this sequence is more prevalent or typical than
CC others, all of which have been entered in this library with
CC annotation. The HTLV-III genome encodes at least six proteins or
CC polyproteins: gag, pol, env, TAT, 27K antigen and the so-called 23K
CC product. The 3' ORF (positions 8797-9447) is truncated in BH10
CC (stop codon at positions 8186-8198), but reads through in BH8 and
CC other sequences to yield what is now called the 27K antigen. The
CC sequence below is from BH10 with exception of the variation at
CC position 9197 which allows annotation of the 27K coding sequence.
Additionally there are four short open reading frames, bases

CC 10932-10935, 43412-46412, 5532-5826 and 6035-6340, which are conserved
 CC to a large degree. A seventh gene has been proposed based upon a
 CC combination of mutational and regulatory evidence: called "ART" (for
 CC anti-repression transactivator), its product appears to act
 CC post-transcriptionally to relieve negative repression of gag and
 CC env production [21]. The exon assignments for ART are putative, but
 CC if they are corroborated, the ART protein would be 116 amino acids
 CC in length. The mechanism for pol gene translation has not been
 CC elucidated; a gag-pol fusion protein is possible; splicing or
 CC frameshift have not been ruled out. The viral protease would be
 CC determined by the region in question. Approximately two-thirds of
 CC the variant sites in the gag and pol genes are "silent mutations",
 CC while over half of those in the env gene are not. Reference [20]
 CC defines divergent and conserved regions for the env gene. Because
 CC of the excessive variability of the env gene, differences between
 CC the sequences summarized herein and other env gene entries have not
 CC been enunciated; only HTLV-III sequence variations have been
 CC included in the sites of this entry. Other entries will include
 CC information for alignment with this entry, including the Zaire and
 CC New York isolate sequences reported by [20]. The TAT protein
 CC (trans-activator protein, approximately 14 kd) is an effector of an
 CC autostimulatory pathway through interaction with a positive control
 CC element, the trans-activating responsive sequence, TAR. TAT seems
 CC to be a transcriptional control molecule in HTLV-I, but [11]
 CC demonstrates that it is a post-transcriptional regulatory molecule
 CC in HTLV-III. Deletion mutants in the TAT gene are incapable of
 CC prolific replication and exhibit no cytopathic effects in T4+ cell
 CC lines [18]. The TAR sequence(s) are found to be between -17 and +80
 CC relative to the cap site +1 (base 455) and is highly conserved.
 CC Enhancer sequences which need not be viral-specific are found
 CC upstream from TAR [8] [11]. Three tandem decanucleotide Sp1 binding
 CC sites are located between bases 377 and 409, of which site III
 CC shows the strongest affinity for the cellular factor; intact, the
 CC three sites cause up to a tenfold effect on transcriptional
 CC efficiency in vitro [17] (The authors demonstrate the existence of
 CC Sp1 in a human T-cell line). In addition to the ~9.4 kb genomic
 CC mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 have been
 CC detected. All are probably polyadenylated at the same site,
 CC position 8665 below, with a potential polyadenylation signal at
 CC 8042-8046 and capped at the same site, position 455, with a
 CC potential TATA box at 427-431. The doubly-spliced transcript of
 CC about 4.6 kb is responsible for the TAT message at least, and
 CC depending upon the acceptor site, also for the sor and 27K
 CC messages, given that a single, albeit partial, mRNA exists for all
 CC three [18]. The acceptor splice for TAT is at position 5811 and the
 CC putative acceptor splice for 27K is at position 6010; the donor
 CC splice site in all three cases would be at position 6079 [18]. The
 CC doubly spliced message would also encode the newly proposed ART
 protein.

XX

FH	Key	From	To	Description
FT	RPT	1	634	5' LTR
FT	RPT	1	634	5' LTR
FT	VARIANT	82	82	a in BH10; g in H9
FT	VARIANT	101	101	g in BH10; a in H9
FT	VARIANT	108	108	a in [2], H9; g in HXB2 [1]
FT	VARIANT	164	164	g in [2]; t in HXB2 [1], H9
FT	VARIANT	166	166	t in [2]; g in HXB2 [1], H9
FT	VARIANT	176	176	a in [2]; g in HXB2 [1], H9
FT	VARIANT	183	183	c in [2], H9; t in HXB2 [1]
FT	VARIANT	227	227	a in [2], H9; g in HXB2 [1]
FT	VARIANT	281	281	a in [2]; g in HXB2 [1], H9
FT	VARIANT	333	333	c in [2]; t in HXB2 [1], H9
FT	SITE	377	388	Sp1 binding site III [17]
FT	SITE	389	392	Sp1 binding site II [17]

FT	SITE	389	403	Sp1 binding site I [17]
FT	VARIANT	421	421	c in BH10, BH5; t in H9
FT	RPT	454	551	R repeat 5' copy
FT	RPT	454	551	R repeat 5' copy
FT	PROVRL	454	9668	HTLV3 virion RNA
FT	CAP	455	455	genomic mRNA start (cap site) [10]
FT	CAP	455	455	TAT, ART mRNA exon 1 start (cap site)
FT				[10], [18], [21]
FT	VARIANT	501	501	a in BH10, BH5, H9; g in HXB2 [1]
FT	SITE	636	653	primer (Lys-tRNA) binding site
FT	VARIANT	654	654	c in BH10, BH5; t in H9
FT	VARIANT	677	677	g in BH10, BH5; ggag in H9
FT	VARIANT	704	704	tgg in BH10, H9; g in BH5 [2]
FT	CDS	787	2325	gag polyprotein precursor
FT	VARIANT	1250	1230	a in BH10; g in BH5 [2], H9
FT	VARIANT	1431	1431	a in BH10; g in BH5 [2], H9
FT	VARIANT	1455	1455	t in BH10, H9; c in BH5 [2]
FT	VARIANT	1611	1611	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1620	1620	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1650	1650	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1682	1682	t in BH10; c in BH5 [2], H9
FT	VARIANT	1675	1675	g in BH10, BH5; c in H9
FT	VARIANT	1722	1722	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1805	1805	g in BH10, BH5; a in H9
FT	VARIANT	1845	1845	a in BH10, BH5; g in H9
FT	VARIANT	1903	1903	a in BH10, H9; t in BH5 [2]
FT	VARIANT	1906	1906	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1923	1923	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1950	1950	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1953	1953	g in BH10, H9; t in BH5 [2]
FT	VARIANT	1983	1988	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1992	1932	c in BH10, H9; a in BH5 [2]
FT	VARIANT	2003	2003	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2013	2013	g in BH10, H9; a in BH5 [2]
FT	CDS	2391	5129	pol polyprotein (NH2-terminus uncertain; AA at 2391)
FT	VARIANT	2468	2468	g in BH10, BH5; a in H9
FT	VARIANT	2581	2581	c in BH10, H9; t in BH5 [2]
FT	VARIANT	2600	2600	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2741	2741	g in BH10; a in BH5 [2], H9
FT	VARIANT	2827	2827	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2855	2855	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2950	2980	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3007	3007	tta in BH10, H9; gtg in BH5 [2]
FT	VARIANT	3057	3057	a in BH10; g in BH5 [2], H9
FT	VARIANT	3122	3122	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3232	3222	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3302	3302	ag in BH10, H9; ga in BH5 [2]
FT	VARIANT	3368	3368	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3383	3383	g in BH10, BH5; a in H9
FT	VARIANT	3385	3385	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3755	3755	a in BH10, BH5; g in H9
FT	VARIANT	3767	3767	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3833	3833	t in BH10, BH5; c in H9
FT	VARIANT	3855	3855	t in BH10, BH5; c in H9
FT	VARIANT	3858	3858	c in BH10, BH5; t in H9
FT	VARIANT	3872	3872	a in BH10, H9; g in BH5 [2]
FT	VARIANT	3874	3874	a in BH10, BH5; g in H9
FT	VARIANT	3874	3874	g in BH10, BH5; c in H
FT	VARIANT	3872	3962	caa in BH10, H9; tag in BH5 [2]
FT	VARIANT	3877	3877	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3874	4684	c in BH10, H9; a in BH5 [2]
FT	VARIANT	3893	3893	a in BH10, H9; c in BH5 [2]
FT	VARIANT	4010	4010	a in BH10; g in BH5 [2], H9
FT	VARIANT	4016	4016	g in BH10, H9; a in BH5 [2]
FT	VARIANT	4029	4029	t in BH10, H9; c in BH5 [2]

FT	VARIANT	4040	4048	a in BH10; g in BH5 [2], H9
FT	VARIANT	4064	4064	c in BH10, H9; t in BH5 [2]
FT	VARIANT	4116	4116	a in BH10, BH5; c in H9
FT	VARIANT	4167	4167	g in BH10, BH5; c in H9
FT	VARIANT	4282	4282	t in BH10, H9; a in BH5 [2]
FT	CDS	5076	5052	sor 23K protein
FT	VARIANT	5158	5158	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5314	5314	t in BH10, BH5; c in H9
FT	VARIANT	5340	5348	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5401	5401	t in BH10, H9; c in BH5 [2]
FT	VARIANT	5412	5412	c in BH10, H9; t in BH5 [2]
FT	VARIANT	5540	5548	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5628	5628	g in BH10, H9; a in BH5 [2]
FT	VARIANT	5646	5646	g in BH10, H9, HXB3; a in BH5 [2]
FT	CDS	5864	5078	TAT protein, exon 2 (first expressed exon)
FT	VARIANT	5834	5934	a in BH10, H9, HXB3; c in BH5 [2]
FT	CDS	6003	6078	ART protein, exon 2 (first expressed exon); putative
FT	VARIANT	6036	6045	cctcctcaagg in BH10, HXB3 [7]; gctcatcgaa in BH8 [2]; g in BH5 [2], clone 12 cDNA [2]
FT	VARIANT	6086	6086	g in BH10, BH8, H9; a in HXB3 [7]
FT	VARIANT	6096	6096	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6108	6108	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6113	6114	gc in BH10, HXB3 [7], H9; gtaac in BH8 [2]
FT	VARIANT	6124	6124	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6152	6152	g in BH10, HXB3 [7], BH8; c in H9
FT	CDS	6256	6825	envelope protein precursor (env)
FT	VARIANT	6373	6373	a in BH10, HXB3 [7], H9; t in BH8 [2]
FT	VARIANT	6474	6474	t in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	6748	6748	t in BH10, HXB3 [7], H9; a in BH8 [2]
FT	VARIANT	6829	6929	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	7038	7038	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7119	7119	a in BH10, HXB3 [7], H9; g in BH8 [2]
FT	VARIANT	7121	7123	cca in BH10, H9; cac in BH8 [2], HXB3 [7]
FT	VARIANT	7171	7172	gt in BH10, H9; aa in BH8 [2], HXB3 [7]
FT	VARIANT	7187	7187	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7272	7273	aa in BH10, H9; gc in BH8 [2], HXB3 [7]
FT	VARIANT	7291	7291	a in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	7343	7343	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7430	7454	gtttaatagtacttgg in BH10, HXB3 [7], and H9
FT	VARIANT	7461	7481	a in BH10, BH8 [2]; g in HXB3 [7], H9
FT	VARIANT	7481	7493	c in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7521	7521	a in BH10, BH8 [2]; t in HXB3 [7], H9
FT	VARIANT	7574	7574	t in BH10, CH8 [2]; c in HXB3 [7], H9
FT	VARIANT	7636	7636	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7636	7637	cg in BH10, HXB3 [7], H9; gc in BH8 [2]
FT	VARIANT	7645	7645	a in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8060	8061	ca in BH10, BH8 [2], H9; ac in H
FT	VARIANT	8127	8127	a in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	8131	8131	t in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	8135	8135	c in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8257	8257	g in BH10, BH8, HXB3; a in H9
FT	VARIANT	8273	8273	t in BH10, BH8, HXB3; g in H9
FT	VARIANT	8364	8364	g in BH10, HXB3 [7]; a in BH8 [2], H9
FT	CDS	8408	8454	TAT protein, exon 3 (AA at 8410)
FT	CDS	8409	8623	ART protein, exon 3 (putative; AA at 8411)
FT	VARIANT	8422	8422	t in BH10, HXB3 [7], clone 12 cDNA [21]; a in BH8 [2]; c in H9
FT	VARIANT	8464	9464	g in BH10, BH8, HXB3, clone 12 cDNA [21]; a in H9

FT	VARIANT	8857	8857	g in BH10, BH8 [2]; a in HXB3 [7], H9, clone 12 cDNA [21]
FT	VARIANT	8872	8872	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT	VARIANT	8892	8892	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT	VARIANT	8748	8748	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8758	8758	g in BH10, H9; c in BH8 [2]; a in HXB3 [7], clone 12 cDNA [21]
FT	VARIANT	8771	8771	t in BH10, HXB3 [7], clone 12 cDNA [21], H9; c in BH8 [2]
FT	CDS	9827	9447	27K protein, exon 3 (first expressed exon)
FT	VARIANT	8857	8857	g in BH10, BH8, HXB3, clone 12 cDNA [21]; a in H9
FT	VARIANT	8924	8924	c in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8967	8967	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8979	8979	a in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT	VARIANT	8985	8985	t in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT	VARIANT	8997	8997	a in BH10, BH8; c in H9, clone 12 cDNA [21]
FT	VARIANT	8994	8994	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	9019	9019	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT	RPT	9116	9748	3' LTR
FT	VARIANT	9169	9196	t in BH10, clone 12 cDNA [21]; c in BH8 [2]
FT	VARIANT	9197	9197	g in BH8 [2], H9, clone 12 cDNA [21]; a in BH10 [2]
FT	VARIANT	9216	9216	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT	VARIANT	9222	9223	ga in BH10, clone 12 cDNA [21], H9; ag in BH8[2]
FT	VARIANT	9279	9279	g in BH10, BH8; clone 12 cDNA [21]; t in H9
FT	VARIANT	9283	9283	t in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	VARIANT	9284	9284	t in BH10, H9, clone 12 cDNA [21]; a in BH8 [2]
FT	VARIANT	9291	9291	a in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	VARIANT	9297	9297	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	9354	9354	g in BH10, HIVDSM), H9; t in BH8 [2]
FT	VARIANT	9406	9406	a in BH10, BH8; g in H9, clone 12 cDNA [21]
FT	VARIANT	9444	9448	c in BH10; t in BH8 [2], H9, clone 12 cDNA
FT	VARIANT	9536	9563	c in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	RPT	9570	9556	R repeat 3' copy
FT	VARIANT	9616	9616	g in HXB2; a in H9, clone 12 cDNA [21]
FT	VARIANT	9621	9621	g in HXB2; a in H9, clone 12 cDNA [21]
FT	VARIANT	9663	9683	t in BH10, H9; tg in clone 12 cDNA [21]
FT	POLYA	9666	9656	TAT, ART, 27K mRNA exon 3 end (poly-A site) [10],[18],[21]
FT	POLYA	9696	9666	genomic mRNA end (poly-A site) [10]
XX				
SP	Sequence	8740	BP	3031
SP				31791
SP				G: 2369
SP				A: 2168
SP				T: 2168
SP				C: 2168

Initial Score = 1872 Optimized Score = 2186 Significance = 0.00
Residue Identity = 89% Matches = 2255 Mismatches = 157
Gaps = 110 Conservative Substitutions = 0

X	1.0	20	30	40	50	60
AAGAG-CAG---AAGACAGTGGCAATGAGAGTGAAAGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA						
AATAGACAGGTTAATTGATAGACTAATAGAAAG-ACCAGAAGACAGTGGCAAT----GAGAGTGAAGGGAGAA						
X 6210	6220	6230	6240	6250	6260	6270

70	80	90	100	110	120	
AATGGGGCAC-CATGCTCCTTGGATATTGATG-----			AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT			
ATATCAGCACTTGTGCAGATGGGGTGGAGATEGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGC						
6280	6290	6300	6310	6320	6330	6340

130	140	150	160	170	180	
CACAGTCTATTATGGGTAC	-----CT-----GTGTGGAA	-----GGAAGCAA-CCACCA-CTCTATTGGTG				
TACAGAAAAATTGTGGTCACAGTCTATTATGGGTACCTGTGTGGAAAGGAAGCAACCACCACTCTATTGG						
6350	6360	6370	6380	6390	6400	6410

190	200	210	220	230	240
CATCAGATECTAAAGCATATGATA		CAGAGG-TACATA	AT	GTTTGGGCCACACATGCCTG	T
.....
-----TGC	-----ATCAGATECTAAAGCATATGATA	CACAGAGGTACATA	ATGTTGGGC	CACA	CATGCCT
6420	6430	6440	6450	6460	6470

320 330 340 350 360 370
 -ATGACATGGTAAAGAACAGATG-CATGAGGGATATAATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-T
 :
 TTTAACAT-GTGAAA-AAATGACAT---GGTAGAA-CAGATGCATGAGGGATATAA---TCAGTTTATGGGAT
 6540 6550 6560 6570 6580 6590

380	390	400	410	420	430	440
AAAA--TTAACCCCCACTCTGTAGTTAA-AGTGCACTGATTGG---GGAATGCTACTAAT---ACCAA						
CAAACCTAAACCCA---TGTGTAAAATTAAACCCCCACTCTGTGTTAGTTAAAGTGC-ACTGATTGAAAGAA						
6500	6610	6620	6630	6640	6650	6660

450	460	470	480	490	500	510
TAC	AGTAATAACCAATAGTAGTACCGGGGAGAAATGATGATGGAGAAAGGAGAGATAAAAAACTGCTCTTC					
TEA	TAATCAACCAATAGTAGTACCGGGGAGAAATGATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTC					
6670	6680	6690	6700	6710	6720	6730

520 530 540 550 560 570 580
 TATCAGCACAAAGATAAGAGGTAAGGTCTCAGAAAAGAATATGCATTTTTTATAAACTTGATATAATACCAAT
 ||||||| :
 TATCAGCACAAAGATAAGAGGTAAGGTCTCAGAAAAGAATATGCATTTTTTATAAACTTGATATAATACCAAT
 5740 5750 5760 5770 5780 5790 5800

580	5800	610	620	630	640	650
AGATATACTATACTACCGAGCTATACTTCAACAGTTGTAACACCTCAGTCATTACACAGGGCTGTCCAAAGGT						
AGATATACTATACTACCGCTATACTTCAACAGTTGTAACACCTCAGTCATTACACAGGGCTGTCCAAAGGT						
6810	6820	6830	6840	6850	6860	6870

ATCCTTCAACCAATTCCCATACATTATTGTGCCCGGCTGGTTTGCATTCTAAAATGTAATAAGAC
ATCCTTCAACCAATTCCCATACATTATTGTGCCCGGCTGGTTTGCATTCTAAAATGTAATAAGAC
6680 6680 6900 6910 6920 6930 6940
740 750 760 770 780 790 800
GTTCAATGGAACAGBACCATGTACAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC
GTTCAATGGAACAGBACCATGTACAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC
6950 6960 6970 6980 6990 7000 7010 7020
810 820 830 840 850 860 870
AACTCAACTGCTATTCAATGGCACTCTAGCAGAAGAAGAGGTTAGTAATTAGATCTGCCAATTCACAGACAA
AACTCAACTGCTGTTAAATGGCACTCTAGCAGAAGAAGAGGTTAGTAATTAGATCTGCCAATTCACAGACAA
7030 7040 7050 7060 7070 7080 7090
880 890 900 910 920 930 940
TGCTYAAACCTAATAGTACAGGCTGAAACCAATCTGTAGAAATTAATTGTACAAAGACCCAAACAACAAATACAAG
TGCTYAAACCTAATAGTACAGGCTGAAACCAATCTGTAGAAATTAATTGTACAAAGACCCAAACAACAAATACAAG
7100 7110 7120 7130 7140 7150 7160
950 960 970 980 990 1000 1010
AAAAAGTATCCTATCCAGAGGACCCAGGGAGAGCATTGTTACAATAGGAAAAATAGGAAATATGAGACA
AAAAAGTATCCTATCCAGAGGACCCAGGGAGAGCATTGTTACAATAGGAAAAATAGGAAATATGAGACA
7170 7180 7190 7200 7210 7220 7230
1020 1030 1040 1050 1060 1070 1080 1090
AGCACATTGTAACATTAGTAGACCAAAATGCAATGCCACTTTAACAGATAGCTAGCAAATTAAGAGAAC
AGCACATTGTAACATTAGTAGACCAAAATGCAATGCCACTTTAACAGATAGCTAGCAAATTAAGAGAAC
7240 7250 7260 7270 7280 7290 7300
1100 1110 1120 1130 1140 1150 1160
ATTTGGAAATTAATAAACAAATACTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAACGCACAGTTT
ATTTGGAAATTAATAAACAAATACTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAACGCACAGTTT
7310 7320 7330 7340 7350 7360 7370 7380
1170 1180 1190 1200 1210 1220 1230
TAAATGTCAGGGAAATTTCTACTGTAATTCAACACAACTGTTAACAGTACTTGGTTAACAGTACTTG
TAAATGTCAGGGAAATTTCTACTGTAATTCAACACAACTGTTAACAGTACTTGGTTAACAGTACTTG
7390 7400 7410 7420 7430 7440 7450
1240 1250 1260 1270 1280 1290 1300
GAGTACTGAAAGCTAAACAACTGAAAGGAAGTGACACAATCACACTCCCAGCAGAATAAAACAATTAT
GAGTACTAAAGGCTAAACAACTGAAAGGAAGTGACACAATCACCTCCAGCAGAATAAAACAATTAT
7460 7470 7480 7490 7500 7510 7520
1310 1320 1330 1340 1350 1360 1370
AAACATGTCAGGGAAATGAGGAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA
AAACATGTCAGGGAAATGAGGAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA
7530 7540 7550 7560 7570 7580 7590
1380 1390 1400 1410 1420 1430 1440 1450
TATTACAGGGCTACTTAACAAACAGATGGTGTTAATAACAAACAATGGGTCCGAGATCTTCAGACCTGGAGG
TATTACAGGGCTACTTAACAAACAGATGGTGTTAATAACAAACAATGGGTCCGAGATCTTCAGACCTGGAGG
7600 7610 7620 7630 7640 7650 7660

AGGAGATAATGGGGAAUATTGAAAGTGAATTATAAATATAAAAGTAGTAAAAATTGAACCATTAGGAGT
7670 7680 7690 7700 7710 7720 7730 7740

1530 1540 1550 1560 1570 1580 1590
AGCAGCCACCAAAAGCTTGTTCCT
1590 1600 1610 1620 1630 1640 1650 1660
AGCAGCCACCAAAAGCTTGTTCCT
7750 7760 7770 7780 7790 7800 7810

1600 1610 1620 1630 1640 1650 1660
TGGGTTCTTGGGAGCGAGGAAACACTATGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
1660 1670 1680 1690 1700 1710 1720 1730
TGGGTTCTTGGGAGCGAGGAAACACTATGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
7820 7830 7840 7850 7860 7870 7880

1670 1680 1690 1700 1710 1720 1730
ATTGCTCTGGTATAGTCACCGAGAGAACAAATTGCTGAGGGCTATTGAGGCAGAACAGCATCTGTTGCAACT
1730 1740 1750 1760 1770 1780 1790 1800 1810
ATTGCTCTGGTATAGTCACCGAGAGAACAAATTGCTGAGGGCTATTGAGGCAGAACAGCATCTGTTGCAACT
7890 7900 7910 7920 7930 7940 7950

1740 1750 1760 1770 1780 1790 1800 1810
CACAGTCTGGGAGCTGAGGCTGAGGCAAGAACATCCTGGCTGTGGAAAGATACTAAAGGATCAACAGCT
1810 1820 1830 1840 1850 1860 1870 1880
CACAGTCTGGGAGCTGAGGCTGAGGCAAGAACATCCTGGCTGTGGAAAGATACTAAAGGATCAACAGCT
7960 7970 7980 7990 8000 8010 8020

1820 1830 1840 1850 1860 1870 1880
CCTGGGAAATTGGGTTGCTCTGAAAACCTCATTGACCAACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAA
1880 1890 1900 1910 1920 1930 1940 1950
CCTGGGAAATTGGGTTGCTCTGAAAACCTCATTGACCAACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAA
8030 8040 8050 8060 8070 8080 8090 8100

1890 1900 1910 1920 1930 1940 1950
TAATCTCTGGAAATACATGACCTGGATGGAGTGAGGACAGAGAAATTAAACAATTACACAAG
1950 1960 1970 1980 1990 2000 2010 2020
TAATCTCTGGAAATACATGACCTGGATGGAGTGAGGACAGAGAAATTAAACAATTACACAAG
8110 8120 8130 8140 8150 8160 8170

1960 1970 1980 1990 2000 2010 2020
CTTAACTACCTTAAATTGAAATGAAATCGCAAAACAGCAAGAGAAATGAACAAGAATTATTGGAATTAGA
2020 2030 2040 2050 2060 2070 2080 2090
CTTAACTACCTTAAATTGAAATGAAATCGCAAAACAGCAAGAGAAATGAACAAGAATTATTGGAATTAGA
8180 8190 8200 8210 8220 8230 8240

2030 2040 2050 2060 2070 2080 2090
TAATGGGCAACTTTGGATTGTTAACATAACAAATTGGCTGTGGTATATAAAATTTCATAATGAT
2090 2100 2110 2120 2130 2140 2150 2160 2170
TAATGGGCAACTTTGGATTGTTAACATAACAAATTGGCTGTGGTATATAAAATTTCATAATGAT
8250 8260 8270 8280 8290 8300 8310

2100 2110 2120 2130 2140 2150 2160 2170
AGTAAAGAGAGTTGTTAGTAAAGTTTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA
2170 2180 2190 2200 2210 2220 2230 2240
AGTAAAGAGAGTTGTTAGTAAAGTTTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA
8320 8330 8340 8350 8360 8370 8380

2180 2190 2200 2210 2220 2230 2240
TTCACCATTTGGTCAAGACCCACCTCCAAACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
2240 2250 2260 2270 2280 2290 2300 2310
TTCACCATTTGGTCAAGACCCACCTCCAAACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
8390 8400 8410 8420 8430 8440 8450 8460

AGGTGAGAGAGAGAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT
 AGGTGAGAGAGAGAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT
 8470 8480 8490 8500 8510 8520 8530
 2390 2380 2340 2350 2360 2370 2380
 GCCTAGGCGTTTCAGCTTCAAGCTACCAACCGCTTGAGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC
 GCCTAGGCG-TTCAGCTTCAAGCTACCAACCGCTTGAGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC
 8540 8550 8560 8570 8580 8590 8600
 2380 2400 2410 2420 2430 2440 2450
 TTCTGGGACGCCAGGGGTTGGGAAAGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACTAA
 TTCTGGGACGCCAGGGGTTGGGAAAGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAGCTAA
 8610 8620 8630 8640 8650 8660 8670
 2460
 AG
 ||
 AG
 X

3. KUNZ-158-CL32. SEQ

HIVM11.CDG Human lymphadenopathy virus (MAL isolate), comple

ID HIVM11.CDG standard RNA: 9229 BP.
 XX
 AC X04415.
 XX
 DT 17-OCT-1986 (incorporated)
 XX
 DE Human lymphadenopathy virus (MAL isolate), complete genome.
 XX
 KW acquired immune deficiency syndrome; env gene; gag gene; genome;
 KW long terminal repeat; pol gene; polyprotein; provirus;
 KW reverse transcriptase.
 XX
 DS Human lymphadenopathy virus
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-9229)
 RA Alizon M., Wain-Hobson S., Montagnier L., Sonigo P.;
 RT "Genetic variability of the AIDS virus: Nucleotide sequence
 RT analysis of two isolates from African patients";
 RL Cell 46:63-74(1986).
 XX
 CC Acquired immune deficiency syndrome (AIDS) is caused by a
 CC retrovirus known by several different names, probably representing
 CC two separate strains: human T-cell lymphotropic virus-III
 CC (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought
 CC to be one strain; and AIDS-associated retrovirus type 2 (ARV-2) the
 CC other. All three viruses, whose sequences do not differ by more
 CC than about 5%, are believed to belong to the retroviral subfamily
 CC Lentiviridae, or "slow" viruses. For the details of the annotation
 CC and for other pertinent references, see the HIV reference entry.
 XX

FH	Key	From	To	Description
FT	RPT	1	96	R repeat 5' copy
FT	RPT	1	177	5' LTR
FT	S7NE	179	186	primer (Lys-tRNA) binding site
FT	GUS	350	1367	gag polyprotein
FT	GUS	1863	4571	pol polyprotein (NH2-terminus uncertain: AA at 1963)

FT 008 4616 5154 ssp 23K protein
 FT 013 5134 5424 urfC
 FT 016 5405 5619 tat protein, exon 2 (first expressed exon)
 FT 016 5729 8376 envelope polyprotein precursor
 FT 016 7869 8007 tat protein, exon 3 (AA at 7960)
 FT 016 8390 9008 27K protein
 FT RPT 8676 8229 3' LTR
 FT RPT 8134 5229 R repeat 3' copy
 XX
 SQ Sequence 5229 BP: 3355 At: 1627 C: 2204 G: 2043 T: 0 other:
 Initial Score = 316 Optimized Score = 2060 Significance = 0.00
 Residue Identity = 84% Matches = 2085 Mismatches = 350
 Gaps = 43 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
AAGAAGCAAGAAGACAGTGGCAATGAGTAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAAATGGG							
AAGAAGCAAGAAGATAGTGCGAATGAGAGTGAGGGAGATAACA-GAGGAATTATCAAAA--CTGGTGGAGATGGG							
5780	5790	5800	5810	5820	5830	5840	

80	90	100	110	120	130	140	
GCACCATECTCCCTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG							
GCATGATCTCTCTGGGATTTGATGACCTGTAGTATTGCAGAAGATTGTGGGTACAGTTATTATGGGG							
5850	5860	5870	5880	5890	5900	5910	

150	160	170	180	190	200	210	
TACCTGTGAGGAAAGCAACCAACTCTATTTCGATCATCAGATGCTAAAGCATATGATAACAGAGGTAC							
TACCTGTGAGGAAAGCAACCAACTCTATTTCGATCATCAGATGCTAAATCATATGAAACAGAAGTAC							
5920	5930	5940	5950	5960	5970	5980	5990

220	230	240	250	260	270	280	
ATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA							
ATAACATCTGGGATACACATGCCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATGTCA							
6000	6010	6020	6030	6040	6050	6060	

290	300	310	320	330	340	350	360
CAGAATTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTATGGGATC							
CAGAAGGGTTAACATGTGGAAAAATAACATGGTAGCAGATGCATGAGGATATAATCAGTTATGGGATC							
6070	6080	6090	6100	6110	6120	6130	

370	380	390	400	410	420		
AAACCTAAAGGCGATCTTAAATTAAACCCACTCTGTGTAGTTAAAGTGCAC TGATTG---GGGAATG							
AAACCTAAAGGCGATCTTAAACGTAACCCACTCTGTGTCACTTTAAACTGCACATGTGAATGGGACTG							
6140	6150	6160	6170	6180	6190	6200	

430	440	450	460	470	480	490	
CTAATAT---ACCAATACTAGTAATACCAATAGTAGTAGCGGGGAAATGATGATGGAGAAAGGAGAGATAA							
CTTGATTAATGACTTAATGCTGGGAA---AATAGGACTAATGCAGAATTGAAAATGGGAGAAGTGA							
6210	6220	6230	6240	6250	6260	6270	

500	510	520	530	540	550	560	570
AAAAGCTGCTCTTCAATATCACGACAAUNATAAGACGGTAAGGGTGCAAGAAAGAATATGCATTTTTTATAAAC							
AAAAGCTGCTCTTCAATATAACCCAGTAGGAAGTGATAAAAGGC---AAGAATATGCACACTTTTATAACC							
6280	6290	6300	6310	6320	6330	6340	

580	590	600	610	620	630
TTGGATATATGGGAA---GTTGATGAGCTACGACCTATACGTTGACAAAGTTGTAAACGACCTCAGTCATTA					

TTGATCTAGTACAAATAGATGATACTGATAAAGCTATAGGCTAATAATTGTAATACCTCAGTAATTA
6360 6360 6370 6380 6390 6400 6410
640 650 660 670 680 690 700 710
CACAGGGCTTCGAAAGGTATGTTTGAGGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTGCGATTG
6420 6430 6440 6450 6460 6470 6480
720 730 740 750 760 770 780
TAAYATGTAATAATAAGACCTTCATGAAACAGGACCAGTACAAATGTCAGCACAGTACAATGTACACATG
6490 6500 6510 6520 6530 6540 6550 6560
750 800 810 820 830 840 850
GAATTAAGCCAGTAGTTCATCACTGAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGGTAGTAATTAGAT
6570 6580 6590 6600 6610 6620 6630
860 870 880 890 900 910 920
CTGCGAATTCGAGAACATGCTAAACCCATAATAGTACAGCTGAACCAATCTGTAGAAATTAAATTGTACAA
6640 6650 6660 6670 6680 6690 6700
930 940 950 960 970 980 990
GACCGAAACAAATAGAAGAAAGGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTGTTACAATAGGAA
6710 6720 6730 6740 6750 6760 6770
1000 1010 1020 1030 1040 1050 1060 1070
AAATAGGAAATATGAGAAAGCACATTGTAACATTAGTAGAGGCAAAATGCAATGCCACTTTAAACAGATAG
6780 6790 6800 6810 6820 6830 6840
1080 1090 1100 1110 1120 1130 1140
CTAGCAAAATTAGAGAACAAATTGAAATAATAAAACAAATAATCTTAAGCAATCCTCAGGAGGGGACCCAG
6850 6860 6870 6880 6890 6900 6910
1150 1160 1170 1180 1190 1200 1210
AAATGTAACGAGTTTAAATTGTGGAGGGAAATTTCTACTGTAATTCAACACAACTGTTAATAGTA
6920 6930 6940 6950 6960 6970 6980
1220 1230 1240 1250 1260 1270 1280
CTTGCTTAACTTAACTTGGAGCTGAAAGGTCAAATAACACTGAAAGGAAGTGACACAATCACACTCCCAT
6990 7000 7010 7020 7030 7040 7050
1290 1300 1310 1320 1330 1340 1350
GCAGGAAATGAAATGAAATGAAATGTTGGAGGAAGTAGGAAAAGCAATGTATGCCCTCCATCGCGGGAC
7060 7070 7080 7090 7100 7110 7120
1360 1370 1380 1390 1400 1410 1420
GCAGGAAATGAAATGAAATGTTGGAGGAAGCTGAAAGGAAAGCTATGATGCCCCCTCCATCGCAGGAG
1380 1390 1400 1410 1420
GCAGGAAATGAAATGAAATGTTGGAGGAAGCTGAAAGGAAAGCTATGATGCCCCCTCCATCGCAGGAG
1390 1400 1410 1420
GCAGGAAATGAAATGAAATGTTGGAGGAAGCTGAAAGGAAAGCTATGATGCCCCCTCCATCGCAGGAG
1400 1410 1420
GCAGGAAATGAAATGAAATGTTGGAGGAAGCTGAAAGGAAAGCTATGATGCCCCCTCCATCGCAGGAG
1410 1420
GCAGGAAATGAAATGAAATGTTGGAGGAAGCTGAAAGGAAAGCTATGATGCCCCCTCCATCGCAGGAG
1420

TCATCAAGCTGTTTATCGAAATACTTACAGGGCTGATATTAAACAAGAGATGGTGAAATAGTAGTGACAATAGTG
7130 7140 7150 7160 7170 7180 7190

1430 1440 1450 1460 1470 1480 1490
TCAGAGAGATTCAGACCTGGAGGAGATATGGGGACAATTGGAGAAGTGAATTATATAAATATAAAG
ACAGAGAGAGATTCAGACCTGGAGGAGATATGGGGACAATTGGAGAAGTGAATTATATAAATATAAAG
7200 7210 7220 7230 7240 7250 7260

1500 1510 1520 1530 1540 1550 1560
TAATGAAATTTAACATTTCAGCTAGCACCACCCAGGCAAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAG
TAACTGGAAATTTCAGCTAGCACCACCCAGGCAAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAG
7270 7280 7290 7300 7310 7320 7330

1570 1580 1590 1600 1610 1620 1630 1640
CAATGAGAAATAGAGCTTTCCTGGGTTCTGGAGCAGCAGGAAGCAGTATGGGCGCACGGTCAATGA
CAATGAGAAATAGAGCTTTCCTGGGTTCTGGAGCAGCAGGAAGCAGTGGGCGCACGGTCAATGA
7340 7350 7360 7370 7380 7390 7400 7410

1650 1660 1670 1680 1690 1700 1710
CGCTTACGCTTACAGACAGATTATTGTCTGGTATAGTGCAGCAGCAGAACATTTGCTGAGGGCTATTG
CGCTTACGCTTACAGACAGATTATTGTCTGGTATAGTGCAGCAGAACATTTGCTGAGGGCTATTG
7420 7430 7440 7450 7460 7470 7480

1720 1730 1740 1750 1760 1770 1780
AGGAGCAACAGGCTCTTGCAGTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAACATCCTGGCTGTGG
AGGAGCAACAGGCTCTTGCAGTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAACATCCTGGCTGTGG
7530 7540 7550

1790 1800 1810 1820 1830 1840 1850
AAAGATACTTAAAGCTCAACACGCTCTGGGGATTCTGGGGTTGCTCTGGAAAACACTCATTGCACCACTGCTG
AAAGATACTTAAAGCTCAACACGCTCTGGGGATTCTGGGGTTGCTCTGGAAAACACATTGCACCACTTGG
7630 7640 7650 7660 7670 7680 7690

1860 1870 1880 1890 1900 1910 1920
TGCCTTGGAAATCTGAGTATAAAATCTGGAAACAGATTGGAAATAACATGACCTGGATGGAGTGG
TGCCTTGGAAATCTGAGTATAAAATCTGGAAATAATATGACCTGGATGCAGTGG
7630 7640 7650 7660 7670 7680 7690

1930 1940 1950 1960 1970 1980 1990 2000
ACAGAGAAATTAACAACTTACACAGAGCTTAATACATTCTTAATTGAAGAACATCGCAAAACCCAGCAAGAAAAAGA
ACAGAGAAATTAACAACTTACACAGAGCTTAATACAACTTAATTGAAGAACATCGCAAAATCCAGCAAGAAAAAGA
7700 7710 7720 7730 7740 7750 7760 7770

2010 2020 2030 2040 2050 2060 2070
ATGAGCAAAATTTAGAAATGAGCTACATAATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
ATGAGCAAAATTTAGAAATGAGCTACATAATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
7780 7790 7800 7810 7820 7830 7840

2100 2120 2140 2160 2180 2200 2210
GGTAACTTGGAAATCTGAGTATAAAATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
GGTAACTTGGAAATCTGAGTATAAAATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
7850 7860 7870 7880 7890 7900 7910

2110 2130 2150 2170 2190 2200 2210
TAGAGCTGATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
TAGAGCTGATGAGCTGGCAAGTTTGTTAAGCATAACAAATGGCTGT
—AC

TATTAATAGAATTTCGAGGAACTCACCTCTCGTTGCAGACCCCTCCTCCAACACCGAGGGGGACCAC
 7830 7830 7840 7850 7860 7970 7980
 2010 2020 2030 2040 2250 2260 2270 2280
 CCAGCAGGCGTAACTGAACTGAGAGAGAGAGAGACAGAGATCCATTGATTAGTGAAACG
 7990 8000 8010 8020 8030 8040 8050
 2320 2330 2340 2350
 GATTTCTTGTCTTATCTTGACGATCTGGGAGCCTTGTGCCCTTTCAGCTACCACCGCTTGAGAGACTTA
 GATTGTCAGGACTTATCTGGGAGCCTGAGGAACC-TGTGCCCTTCAGTTACCAACCGCTTGAGAGACTTA
 8060 8070 8080 8090 8100 8110 8120
 2360 2370 2380 2390 2400 2410 2420 2430
 CTGTTGATTTGAGGAGATTTTGTAACTTCTGGGACGCCAGGGGGTGGGAAGGCCCTCAAATATTGGTGGAAT
 CTGTTAATTCGAGGAGGATTGTGTAACTTCTGGGACGCCAGGGGGTGGGAAGGCCCTCAAATATCTGTGGAAAT
 8130 8140 8150 8160 8170 8180 8190 8200
 CTCCTACAGGATTTGAGGAGCTAAAG
 CTCCTACAGGATTTGAGGAGCTAAAG
 CTCCTACAGGATTTGAGGAGCTAAAG
 8210 8220 8230

4. KUNST (1950-1955)

HIV-1 Human lymphadenopathy virus (ELI isolate), complet

ID HUMANIMMUNODEFICIENCY VIRUS; 9176 BP.
XX
AC X09K141
XX
DT 17-DEC-91 0000 (Unpublished)
XX
DE Human lymphotropic virus (ELI isolate), complete genome.
XX
KW Acquired immune deficiency syndrome; env gene; gag gene; genome;
KW Human T-cell lymphotropic virus; pol gene; polyprotein; provirus;
KW T-cell lymphotropic viruses.
XX
OS Human lymphotropic virus
OC Viralidae; non-RNA enveloped viruses; Retroviridae.
XX
RN 101 (Cited: 1-1776)
RA ABLER B., BERNSTEIN S., MONTAGNIER L., SONIGO P.;
RT "Genetic variability of the AIDS virus: Nucleotide sequence
analysis of two isolates from African patients";
RL Cell (1985) 41:189-204 (1985).
XX
CC Acquired immune deficiency syndrome (AIDS) is caused by a
CC retrovirus known by several different names, probably representing
CC two separate strains: human T-cell lymphotropic virus-III
CC (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought
CC to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the
CC other. HTLV-III and LAV, whose sequences do not differ by more
CC than about 5%, are believed to belong to the retroviral subfamily
CC of lentiviruses or "slow" viruses. For the details of the annotation
CC and for other pertinent references, see the HIV reference entry.
XX

xx	xx	xx	xx	Description
----	----	----	----	-------------

FH

2110 2120 2130 2140 2150 2160 2170
 CGGCGCTTGTGGATTTGAAATGTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTG
 TTTTCTATAGTGAATAGAGTTAGGCAGGGATATTG
 AGTTCGGATTTGAAATGTTTGCTGTACTTTCTTAGTAAATAGAGTTAGGCAGGGATACTC
 7830 7840 7850 7860 7870 7880 7890
 2180 2190 2200 2210 2220 2230 2240
 ACCGTTATCTTGTGGACGGCCGACAGGCCGAAGGAATAGAAGAAGAAGG
 TTTTCTATAGTGAATAGAGTTAGGCAGGGATATTG
 ACCGTTATCTTGTGGACGGCCGACAGGCCGAAGGAACAGAAGAAGAAGG
 7890 7900 7910 7920 7930 7940 7950 7960 7970
 2250 2260 2270 2280 2290 2300 2310
 TGGACGGATCTGGACGATCTGCG
 TGGACGGATCTGGACGATCTGCG
 TGGACGGATCTGGACGATCTGCG
 7930 7940 8000 8010 8020 8030 8040
 2320 2330 2340 2350 2360 2370 2380
 GACCGCTTGTGGATCTGGACGATCTGCG
 GACCGCTTGTGGACGATCTGCG
 8050 8060 8070 8080 8090 8100 8110
 2390 2400 2410 2420 2430 2440 2450 2460
 TGCG
 TGCG
 TGCG
 8120 8130 8140 8150 8160 8170 8180 X

5. KUNZ-158-CLB1.5B+

REHTLV4C STLV-3 (HTLV-4) partial provirus DNA (lambda PKE10

ID REHTLV4C Standard DNA; 5391 BP.
 XX
 AC M36351 ; Y00265
 XX
 DT 24-MAR-1986 (contamination)
 XX
 DE STLV-3 (HTLV-4) partial provirus DNA (lambda PKE102)
 XX
 KW env gene; gag gene; long terminal repeat; overlapping genes;
 KW unidentified reading frame.
 XX
 OS Human T-cell leukaemia virus type IV
 OC Rhabdoviridae; ds-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-5391)
 RA Hahn B. H.
 RT ;
 RL Submitted (08-DEC-1987) on tape to the EMBL Data Library by:
 RL Hahn B. H., University of Alabama at Birmingham (UAM), 1918
 RL University Blvd, SRSG 288, Birmingham, AL 35295.
 XX
 RN [2]
 RA Hahn B. H., Kong L. T., Lee S. W., Kumar R., Taylor M. E., Arya S. K.,
 RA Shaw G. M.;
 RT "Relation of HTLV-4 to Simian and human
 immunodeficiency-associated viruses";
 RT Nature 330:184-186(1987).
 XX
 CC *source: clone=lambda PKE 102
 CC lambda PKE 102 is derived from PK82(HTLV-IV);
 CC it is assumed that this isolate is a laboratory contaminants

500 570 580 590 600
 GAAATGCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 TTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 600 630 640 650 660 670
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 670 700 710 720 730 740
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 740 770 780 790 800
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 800 830 840 850 860 870
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 870 900 910 920 930 940
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 940 970 980 990 1000 1010
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1010 1040 1050 1060 1070 1080
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1080 1110 1120 1130 1140
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1140 1170 1180 1190 1200
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1200 1230 1240 1250 1260 1270
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1270 1300 1310 1320 1330 1340
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1340 1370 1380 1390 1400 1410
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1410 1440 1450 1460
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1460 1490 1500 1510 1520 1530
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1530 1560 1570 1580 1590 1600
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1600 1630 1640 1650 1660 1670
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1670 1700 1710 1720 1730 1740
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1740 1770 1780 1790 1800
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1800 1830 1840 1850 1860 1870
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1870 1900 1910 1920 1930
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 1930 1960 1970 1980 1990 2000
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 2000 2030 2040 2050 2060 2070
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 2070 2100 2110 2120 2130 2140
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 GCGGAACTTGTATTTTAACTTAACTAATAACCAATAGATAATGATACTAC
 2140

6. 本章小结与思考题

REGISTRATION Human immunodeficiency virus STLV-III(AGM) provir

RL 1740 M 10731 NDDI RING 37 Room 6C09, Bethesda MD 20892, USA.

XX
RN 841 (Chloroform)

RA Pichichet S., Baldo B.C., Guo H.G., Gurgi C., Callatti E.,

RA Margalit R., Y. Hwang, Wong-stael F., Reitz M.S.;

RT "Sequence of simian immunodeficiency virus and its relationship to
the human immunodeficiency viruses";
RL Nature 331:538-543 (1988).

XX
CC Source: library=EMBL-5; cell line=infected K6W

FH	Key	From	To	Description
FT	5'LTR	1	176	R-region of 5'-LTR
FT	5LTR	1	303	5'-long terminal repeat
FT	5'PSE	152	157	put. polyA signal
FT	5'UTR	177	303	U5-region of LTR
FT	T3463	304	323	transfer RNA-Lys(3)
FT	G15	532	923	protein p17 (AA 1 - 115)
FT	G16	538	2057	gag gene product
FT	G17	284	3055	protein p24 (AA 1 - 391)
FT	G18	1714	2265	protease (AA 1 - 184)
FT	G19	1714	4075	pol gene precursor polypeptide
FT	G20T	1725	1756	imp. direct repeat 1
FT	G21T	1708	1819	imp. direct repeat 1
FT	G22S	2266	4575	reverse transcriptase (AA 1 - 870)
FT	G23E	4552	6368	polypurine tract
FT	G24S	4773	6449	sur gene product
FT	G25S	5702	7303	large envelope glycoprotein gp120 (AA 1 - 536)
FT	G26E	5702	8371	env gene product
FT	G27S	7310	8371	small envelope glycoprotein gp32 (AA 1 - 354)
FT	G28	81876	83239	3'-ORF (AA 1 - 211)
FT	G29E	85746	9590	poly purine tract
FT	G30E	85911	9026	U3-region of LTR
FT	G31T	85911	8664	3'-long terminal repeat (LTR)
FT	G32E	8604	8264	R-region of 3'LTR
FT	G33T	8704	1083	put. TATA-box

XX
SQ Sequence: 5264 BP; 3121 A; 1749 C; 2309 G; 2085 T; 0 other;

Initial Score = 293 Optimized Score = 1250 Significance = 0.00
Residue Identity = 55% Matches = 1433 Mismatches = 904
Gaps = 26% Conservative Substitutions = 0

X 10 20 30 40 50 60
AAG-AACCA-TATGCTTG-CATGAGAG-TGAAGGAGAA-ATATCAGCACTT--GTGGAGATGGG---GGTGG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAG-AATGCGATTCG-C-TTGGAGATCAGCTGCTTATCGCCATCT-TGCTTTAAAGTGTCTATGGGATCTATTGT
X 5730 5740 5750 5760 5770 5780 5790

70 80 90 100 110 120 130
A---TATGATGGGATCTTGGGATATTGATGATCTGTAG---TGCTACAGAAAAATTGTGGGTCAACA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AATGCGATTA-TGCA-GATCTT-TTATGGTGTACCGAGCT-TGGAGGAATGCGACAATTCCCCTCTTCTGTGCA
5800 5810 5820 5830 5840 5850 5860

140 150 160 170 180 190
GTCGTTATTGATGGGATCTTGGGATCTTGGGATATTGATGATCTGTAG---TGCTACAGAAAAATTGTGGGTCAACA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ACCGAGAAATGCGATTA-TGCA-GATCTT-TTATGGTGTACCGAGCT-TGGAGGAATGCGACAATTCCCCTCTTCTGTGCA
5870 5880 5890 5900 5910 5920 5930

TTAATGTTTACGAAAGCTTATGGGAGA---AT-ACAGT----CACAGAAC---AGGCAATAGAGG
 5940 5950 5960 5970 5980 5990

200 250 300 310 320 330
 TAAATATGATGAACTATGAAATTAAACATGTGGAAAA--ATGA-CAT--GGTAGAAC---AGAT
 6000 6010 6020 6030 6040 6050 6060
 AGATATGCGGCTTTCAGGAGCCTAAAGGCTTGTAAAATTATCCCATTATGCATTACTATGAGAT

340 350 360 370 380 390
 GCATGAGGATATAATGAGG-TTAT-AGCT---CA-AAGCCTAAA-GCCATGTGTAAAATTAAACCC-AC-T
 6070 6080 6090 6100 6110 6120 6130
 GCATTAACATGAGGAGCATGGATGGGATTCACAAATCATCAACACAATAACAACAGCAGCACCAACAT

400 410 420 430 440 450 460
 CTGTTTAACT-TTAAATGCACTGTTGGGAATGCTACTAATACCAATACTAG-TAATACCAATAGTAGT
 6140 6150 6160 6170 6180 6190 6200
 CACGACCACTATGAAATAATGACATGGTCATGAGACTAGTTCTTGTA-TAGCTCAGAATAATTGCACA

470 480 490 500 510 520 530
 (NEUROBBAV)-ATTAATGATGAGAAAGGAGAGATAAAAAACTGCTCTTCAATATCAGCACAAAGNATAAGAG
 6210 6220 6230 6240 6250 6260 6270
 GCTTGGGAAAGAATGAAATRATAABCTAAATTCAACATGACAGGGTTAAAAGAG-ACAAGACAAAGGA

540 550 560 570 580 590 600
 GTCAGGTTGAAAGAG-AATATGC--ATTT-TTTTATAAACTTGATATAATACCAATAGATAATGATACTAC
 6280 6290 6300 6310 6320 6330 6340
 GTACAAATGAAAGCTTCTTACAGATTGGTTGTGAAC--AAGGGAATAGCACT-GATAATGAAAGCAG

610 620 630 640 650 660 670
 CACGTTACGTTGACAAAGTTGAAACACCTCAGTCATTACACAGGCCGTG-CCAAAGGTATCCTTGAGCCAA
 6350 6360 6370 6380 6390 6400 6410
 ATGATACA---TAAATCACTTAAACACTTCTGTTATCCAAGAGTCTTGTGACAAACAT-TATTGGGATACTA

680 690 700 710 720 730 740
 TTCCGATACATATTGTTGCCCCCGCTGGTTGCGATTCTAAATGTAATAATA-AGACGTTCAATGGAAACA
 6420 6430 6440 6450 6460 6470 6480
 TTAGATTAACTTATTTGTCACCTCAGGTTATGCTTGTCTTAGATGTAATGACACAAATTATTCAAGGCTTTA

750 760 770 780 790 800
 GCAAA-CATGTTACAAATGTCAGCACAGTACAATGTACACATGGA--AT-TAGGC-CAGTAGTATCAACTCA
 6450 6500 6510 6520 6530 6540 6550
 TGCCCTAAATGTTCTAAGGTTGCTCTCTTC-ATG--CACAAAGGATGATGGAGACACAG-ACT-TCTACTTG

810 820 830 840 850 860 870
 ACTGCTGTTGAAATGGCAGTCTACGAGAAAGA-AGAGGGTAGTAATTAGATCTGCCAATTCACAGACAATGCTA
 6580 6590 6600 6610 6620
 GTTTGGCTTTAATGAAACTAAGGAGAAATAGAAACTTATATTAA---CTGGC-ATGGTAGGGATAA---TA

880 890 900 910 920 930 940
 AAACCCATA-TAG---TACAGGCTGAAACCAATCTGTAGAAATTAAATTGTACAAGACCCAAACAACAAATACAAGAA
 6650 6660 6670 6680
 TGGCTATAYATTGTTAAATGAACTTATATAATCTAACAAATGAAATGTAGAAGA-CCAGGA--AAT--AAG--
 950 960 970 980 990 1000 1010
 AGGCTATGTTACGAAAGGAAAGGAAAGGAAAGGAAAGGAAATGAGGAAATATGAG

7440 7450 7460 7470 7480 7490 7500
 1760 1790 1800 1810 1820 1830 1840
 GAATCCTGCGCTGAAAGAGAAGAAGAATTGCGATTGACCGTCTGGGGAAACAAAGAACCTCCAGACTA
 7510 7520 7530 7540 7550 7560 7570 7580
 1850 1860 1870 1880 1890 1900 1910
 TTTCACCAACGCTGCTCTGCTTGCAGT---AATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTTGGAATAACA
 TCTTGACAGCTACTGTACCATGCCAAATGCAAGT---CTAACACCAAGACTGGAACA-AT---GA-TA---
 7590 7600 7610 7620 7630 7640
 1920 1930 1940 1950 1960 1970 1980
 TGACUCTGAACTGAGTGGGACAGAGAAATTAAAC-AATTACACAAAGCTTAATACATTCCCTTAATTGAAGAATCG
 ---CTTGGUAAAGGATGAGGAGAAGGTTGACTTCTTGGAGGAAAATAACA-GCCCTCCCTAGAAGAGGCA
 7650 7660 7670 7680 7690 7700
 1990 2000 2010 2020 2030 2040 2050
 CAAACCAAGAAGAAGAATGAAACAAGAATTATTGGAATTAGATAAAATGGGCAAGTTGTGG-AATTGGTT
 CAAPTTCAAGAAGAAGAACAATGATAATTACAAAAGTTGAATAGCTGGG-ATGTGTTGGCAATTGGTT
 7710 7720 7730 7740 7750 7760 7770
 2060 2070 2080 2090 2100 2110 2120
 TAACATAACAAATTGCGCTGAGTATAAAATATTCA--TAATGATAGTAGGAGGCTTGGTAGGTTTAAGA
 TGACUTTGCGTTGATAAAATATAT-ACAATATGGAATTATG-TAGTTGTAGGAGTAATACTGTTAAGA
 7780 7790 7800 7810 7820 7830 7840
 2130 2140 2150 2160 2170 2180 2190
 ATAGTTTTCGAGTACTTTCATATGAAATAGAGTT-AGGCAGGGATATTACCA---TTATCGTTTCAGAC
 ATAGTGTATATATAGTACAAATGCTGACTA-AGTTAAGGCAGGGGTATAGGCCAGTGTCTC-TTCCCCAC
 7850 7860 7870 7880 7890 7900 7910
 2200 2210 2220 2230 2240
 CCAC---CTCCCA---AC---CCCGAGGGACCCGACAA-GGCC---C---GAAGGAATAGAAGAAGAAG
 CCTTTTATTGCGATGAGCTGATAACCCAACAGRACCCGGCACTGCCAACCAACAGAGAAGGCAAAGAAGGAGACG
 7920 7930 7940 7950 7960 7970 7980 7990
 2250 2260 2270 2280 2290 2300 2310
 GTCGAGAGAGAAGACAGACAGATC---CATTCGATTAGTGAAACGGATCCTTAGCACCTATCTGGGACGAT
 GTCGAGAAACCGTGGCAACAGGCTCCTGGCAGATAGAAATATATTCAATTTC-CTGATCCG---CCAA
 8000 8010 8020 8030 8040 8050
 2320 2330 2340 2350 2360 2370
 CTGGGGAGC---CTTG---TGCGCTTCACTACCUACCGCTTGAGAGACTTACTCTTGATTGTA-ACGAGGAT
 CTGTTACGGCTCTTGACTTGTGCTATTCAAGCAA---CTGCAGAACCTTGCTATCGAGAGCATACCA-GAT
 8060 8070 8080 8090 8100 8110 8120
 2380 2390 2400 2410 2420 2430 2440
 TATGAACTTCGCGAGGCGGAGGCGCT-CAAA-TATTGGTGGAAATCTCCTACAGTATTGGAGT
 CCTGCAACAAATGCTGCGACCCCTACGAAGGGATTGAG--AAGTCCT-CAGGACTGAACT
 8130 8140 8150 8160 8170 8180 8190
 2450 2460
 2470 2480 2490 2500 2510 2520 2530

GACCTTACCTACAA
8200 X

7. KUNZ-158-LTR SEQ
RESIVM0X Simian immunodeficiency virus STLV-III(MAC) provir

ID RESIVM0X standard; DNA; 8646 BP.
XX
AC Y00277
XX
DT 17-SEP-1987 (annotation)
XX
DE Simian immunodeficiency virus STLV-III(MAC) proviral genome
XX
KW art gene; env gene; envelope protein; gag gene; genome;
KW overlapping genes; pol gene; tat gene.
XX
OS Simian immunodeficiency virus
OC Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN 113 (bases 1-8646)
RA Chakrabarti L.
RT
RL Submitted (19-AUG-1987) to the EMBL Data Library.
XX
RN 121 (bases 1-8646)
RA Chakrabarti L., Guyader M., Alizon M., Daniel M. D.,
RA Debrostes R. C., Tchillaïs P., Sonigo P.;
RT "Sequence of simian immunodeficiency virus from macaque and its
RT relationship to other human and simian retroviruses";
RL Nature 328:543-547(1987).
XX
CC #source: library=infected HUT78 cells; clones=lambda SIV1;

FH	Key	From	To	Description
FT	SITE	1	176	LTR (R)
FT	RTT	1	318	5' LTR
FT	SITE	176	318	LTR (U5)
FT	SITE	486	2068	ORF (gag)
FT	CDS	551	2068	gag gene product (AA 1-506)
FT	CDS	1726	4393	pol gene product (AA 1-1056)
FT	SITE	4793	5467	ORF (R)
FT	CDS	4826	5467	R gene product (AA 1-214)
FT	SITE	5293	5633	ORF (X)
FT	CDS	5293	5633	X gene product (AA 1-112)
FT	CDS	5637	5939	R gene product (AA 1-101)
FT	SITE	5758	6083	ORF, exon 1 (tat)
FT	CDS	5783	6083	tat gene product (AA 1-99) (6083 is 2nd base in codon)
FT	SITE	6014	6083	ORF, exon 1 (art)
FT	CDS	6014	6083	art gene product (AA 1-23) (6083 is 1st base in codon)
FT	SITE	6087	6735	ORF (env)
FT	IVS	6084	8300	intron I (tat and art) (8301 is 3rd base in codon)
FT	CDS	8090	8735	envelope protein (AA 1-882)
FT	SITE	8247	9551	ORF, exon 2 (art)
FT	SITE	8251	9394	ORF, exon 2 (tat)
FT	SITE	8298	9300	in-frame stop codon (env)
FT	CDS	9301	9394	tat gene product (AA 100-130)
FT	CDS	8301	9551	art gene product (AA 24-108) (8301 is 2nd base in codon)
FT	SITE	8540	9757	ORF (E)

FT CDS 8572 9357 F-gene product (AA 1-262)
 FT SITE 8557 9171 LTR (U3)
 FT RPT 8557 8545 3' LTR
 FT PRM 9369 9377 enhancer-like sequence
 FT PRM 9382 9391 enhancer-like sequence
 FT SITE 9392 9402 conserved sequence
 FT SITE 9404 9412 conserved sequence
 FT SITE 9414 9424 conserved sequence
 FT SITE 9472 9346 LTR (R)
 FT PRM 9491 9496 pot. promoter sequence
 XX
 SQ Sequence 9646 BP+ 3290 A+ 1817 C+ 2387 G+ 2152 T+ 0 other:
 Initial Score = 250 Optimized Score = 1247 Significance = 0.00
 Residue Identity = 54% Matches = 1433 Mismatches = 897
 Gaps = 281 Conservative Substitutions = 0

X	10	20	30	40	50	60
AAGAAG	-----CAGGAGACAGTGCGAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGA					
ATGGGGTTGTCCTGCGAATCAGCTGC	--TTATCGCCATCTTG-CTTCTAAAGTGTCTATGG-GAT-TTATTGTA					
6090	6100	6110	6120	6130	6140	6150
70	80	90	100	110	120	130
----ANTGGGGCGACCATGC1CCTTGAGATATTGATGATCTGTAG	---TGCTACAGAAAAATTGTGGGTACAG					
TTCAATATGTC-CAGTCT-TTATGGGTGACCGACT-TGGAGGAATGCGACAATTCCCCTCTTCTGTGCAA						
6160	6170	6180	6190	6200	6210	6220
140	150	160	170	180	190	200
TCTATTATGCGGTTACUTGTGTGGAGGAGCCA	--ACC-ACCACTCTATTTG-TGCATCAGATGCTAAAGCA					
CCGAAATATGTCATCT-TGGGGAGAACCTCAAGTGCCTACCAAGATAATGATGATTATTAGA-ATTGGCCCT						
6230	6240	6250	6260	6270	6280	6290
210	220	230	240	250		
T-ATGATACAGAGGTACATAATGTTTG	-----GCCAC---ACATGC--CTGTGTACCCACAG--ACC					
TAATGTTACAGAAAGCTTGTGCTTGGGAGAATACAGTCACAGAACAGGCAATAGAGGACGTATGGCAACT						
6300	6310	6320	6330	6340	6350	6360
260	270	280	290	300	310	320
C----GAAACACAAAGAYAGTAGTATTG-GTAAAATGTGACAGAAAATTTAAC-A-TGTG-GAAAAATGACATGG						
CTTGTGAGACCTCAATAA--AGCCTTGTGTAATATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGT						
6370	6380	6390	6400	6410	6420	6430
330	340	350	360	370	380	390
TAGAACAGATGCATGAGATATAATCAGTTTAT-GGGATCAA-AGCCTAAAGCCATGTGTAATTAACCCC						
GAG-ACAGATTAATGGGGAT-TGACAAAATCATCAACAAACACAGCATCAA--CAACAAACAACAACAGC						
6440	6450	6460	6470	6480	6490	6500
400	410	420	430	440	450	460
ACTGTGTGTTACTTTAAAGTGACTGATTGGGGAAATGCTACTAATACCAATACTAG-TAATACCAATAGTA						
AAAATCAG-TGAGAGAACAGAG-AC--AT--AGTCAATGAGACTAGT-CCTTGTGTAGTTCATGATAATTGCA						
6510	6520	6530	6540	6550	6560	6570
470	480	490	500	510	520	530
GTAGGGGAGA-A-TGA-TGATGAGAGAAAGGAGAGATAAAAAACTGCTCTTCAATATCAGCACAAGNATAAG						
CAGGCTTGTGAGAACAGGCCATGATAAGCTGTAATTCACATGACAGGGTTAAAAAGAG-ACAAGAAAAAG						
6580	6590	6600	6610	6620	6630	6640
540	550	560	570	580	590	

AAGTAAGGTCAGAAAAG--AATATGC--ATTT-TTTTATAAACTTGATATAATACCAATAGATAATGATACT
6650 6660 6670 6680 6690 6700 6710
GATTACAATGAAACTGTGTTGAGATTGGTTGTGAAC--AAGGGAAATAGCACT-GGTAAATGAAAGT
600 610 620 630 640 650 660 670
ACCCGCTAACGTTGACAAGTGTAAACACCTCAGTCATTACACAGGGCTGT-CCAAAGGTATCCTTGAGCC
6720 6730 6740 6750 6760 6770
AGATGTTACA--TAAATCACTGAACTTCTGTTATCCAAGAGTGTGACAAAGAT-TATTGGGATGC
680 690 700 710 720 730 740
AATTCCATACA-TTATTGTCGCCCGGGTGGTTTGCATTCTAAAATGTAATAATA-AGACGTTCAATGGA
TATT-AGATGTAAGATATTGTGACACCTCCAGGTTATGCTTAGATGTAATGACACAAATTATTCAAGGCT
6780 6790 6800 6810 6820 6830 6840
750 760 770 780 790 800
ACACGACCA--TGTACAAATGT-CAGCACAGTACAATGTACACATGGA--AT-TAGGC-CAGTAGTATCAAC
TTATGCCCTAACCTGTTCAAGGTAGTGGTCTCTTC-ATG--CACAAGGATGATGGAGACACAG-ACT-TCTAC
6850 6860 6870 6880 6890 6900 6910
810 820 830 840 850 860 870
TCAACTGCTGTTGAATGGCACTGAGCAGAAGA-AGAGGTAGTAATTAGATCTGCCAATTTCACAGACAAATG
TTGGTTTGGTTTAATGAACTAGAGCAGAAAATAGAACCTATATTAA---CTGGC-ATGGTAGAGATAA--
6920 6930 6940 6950 6960 6970 6980
880 890 900 910 920 930 940
CTAAACCATAA-TGTACACGCTAACCAAT---CTGTAGAAATTAAATTGTACAAGACCCAACAAACAATACA
-TAAGGACTATAATTAGTCTAAAT-AAGCATTATAATCTAACAAATGAAATGTAGAAGA-CCAGGA--AAT--A
6990 7000 7010 7020 7030 7040
950 960 970 980 990 1000 1010
AGAAAAAAATATCOSTATCCAGAGGGACCA--GGGAGAGCATT--TGTACAAATAGGAAAAATAGGAAATAT
AC--ACAGTTT--TA-CCA---GTCACCACTATGTCTGCATTGGTTTCCACT--CACAACCACTCAATGA
7050 7060 7070 7080 7090 7100
1020 1030 1040 1050 1060 1070
GAGAC--AAGCA--CATTTGT-AACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGAT--AGCTAGCA
7110 7120 7130 7140 7150 7160 7170
GAGGCGAAAGCAGGATGGTGTAGGTT--TGGAGGAAATTGGAAGGAGGCAATAAAAGAGGGTGAAGC-AGAC
1080 1090 1100 1110 1120 1130
AATT----AAGAGAAACA--AT-TTGGAAATAATAAAAC-AAT--AATC--TTAAGCAATCCT-CAGGAGG
CATTGTCACACATCCGAGGTAACTGGAACTAACAACTGTATAAAATCAATTGACGGCTCTAGAGGAGG
7180 7190 7200 7210 7220 7230 7240
1140 1150 1160 1170 1180 1190 1200
GGACCCAGAATTAACCCACA-GT--TTAATTGAGGGGAATTGGTACTGTAAATTCAACACAAACT
AGATCCGGAA--GTTACCTGATGTGACAAATTGCAAGAGGAGAGTTCTACTGTAA---AA-TGAATT
7250 7260 7270 7280 7290 7300 7310
1210 1220 1230 1240 1250 1260 1270
GTTTAAATAATACCTGGTTA--ATAGTACTTGGAGTACTGAAGGGGTCAAATAACACTGAAGGGAAAGTGCACACA
GGTTCTA-AATGGGTAGAGATAGGAGTCTAACACTACCCAGAAGCAGAAAGGA-ACGGCATAAAAGG---A
7320 7330 7340 7350 7360 7370 7380
1290 1300 1310 1320 1330 1340

ATGCACTTGGATGCAGATAAAACAATTATTAACATGTGGCAGGAAGTAGGAAAAGCAATG--TATGCC
11 11 1 1 11 11 1 1 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

ATTAC-GTACCGATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAA--AATGTTTATTG
7350 7400 7410 7420 7430 7440

1350 1360 1370 1380 1390 1400 1410
CTCCCATCAGCGAACAAATTAGATGTTCATCAAATAATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATA
11 11 1 1 11 1 1 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
CTCCAAGAGAGGAGAACCTCACCGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAATTGGACTGATG
7450 7460 7470 7480 7490 7500 7510 7520

1420 1430 1440 1450 1460 1470 1480
ACAACAATGGCTCGAG-ATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA--GAAGTGAATTAT
11 11 1 1 11
GAAAUCA---AAGTATBATACT---ACCATGA-GTGCAG--AGGTGGCAGAACTGTATCGATTGGAATTG
7530 7540 7550 7560 7570 7580

1490 1500 1510 1520 1530 1540 1550
ATAAAATATAAACTAAATTGAAACCATTAGGAGTAGCACCCCACCAAGGCAAAGAG---AAGAGTGGT-
11
GAGATTATAAAATTAGAAATCACTCCAATTGGCTTGGCCCCACAAATGTGAAGAGGTACACTACTGGTG
7580 7600 7610 7620 7630 7640 7650

1560 1570 1580 1590 1600 1610 1620
GCA-AGAGAGAAAGAGCAGTGGAAATAGGAGCTTGTCTGGTTCT-TGGGAGCAGCAGGAAGCAC
11
GCAACCTCAAGAAATAAAAG-AG-GGGTCTTGTGCTAGGGTTCTGGGTTTCGCAACGGCAGGTTCTGC
7660 7670 7680 7690 7700 7710 7720

1630 1640 1650 1660 1670 1680 1690
TATGGGCGCAAGGGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTCAGCAGCAGAA
11
AATGGGGCGCGCTCGTTGACCGTACCGCTAGTCCCGGACTTTATTGGCTGGGATAGTCAGCAACAGCA
7730 7740 7750 7760 7770 7780 7790

1700 1710 1720 1730 1740 1750 1760
CAATTTGCCTGAGGCTATTGAGGCCAACAGCATCTTGCACACTCACAGTCTGGGGCATCAAGCAGCTCCA
11
ACAGCTGTGGACGGTGTCAAGAGACAAACAAGAAT1GTTGCGACTGACCGTCTGGGAAACAAAGAACCTCCA
7800 7810 7820 7830 7840 7850 7860

1770 1780 1790 1800 1810 1820 1830
AGCAGAGAATCCTGGCTGTGGAAAGGATACTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAA
11
GACTAGGGITCTGCCATCGAGAAAGTACTTAAAGGACCCAGGCGCAGCTAAATGCTTGGGATGTGCGTTAG
7870 7880 7890 7900 7910 7920 7930

1840 1850 1860 1870 1880 1890 1900
ACTGATTTCACCAACTACTGTGGCTTG--AATGCTAGTTGGAGTAATAAACTCTGGAACAGATTGGAA
11
ACAAGTCTGTACACACTACTGTACCATGGCCAAATGCAAGT---CTAA-----CACCAGA-TTGGAA
7940 7950 7960 7970 7980 7990

1910 1920 1930 1940 1950 1960 1970
TAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACAATTACACAAGCTTAATACA--TTCCTTAATTGAA
11
CAATGAGACTTGGCAAGAGTGGGAGGGAGGGAGGGTTGAC--TTCTTGAGGCAAATATAACGGCCCTCTAGAA
8000 8010 8020 8030 8040 8050 8060

1980 1990 2000 2010 2020 2030 2040
GAATCGCAGAACCAAGCAAGAAAAGAATEAACAGAATTATTGGAATTAGATAAAATGGGCAAGTTGTGG-AA
11
GAGCTACAAATTAACAAAGAGAACAGACATGTATGAAATTACAAAGTGAATAGCTGGG-ATGTGTTGGCAA
8070 8080 8090 8100 8110 8120 8130

2050 2060 2070 2080 2090 2100 2110

TTEGTTAACATAACAAATTGGCTGTGGTATATA-AAAAT---ATTCATAATGATAGTAGGAGGCTGGTAG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TTGGTTGACCTACTTCTGGATAAAGTATATACAATATGGAATTAT-ATAATTGTAGGAG---TAATAC
 8140 8150 8160 8170 8180 8190 8200
 2120 2130 2140 2150 2160 2170 2180
 GTTTAAGAAATAGTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCA---TTATCGT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TGTTAAGAATAGTGAATCTATATAGTACAATGCTAGGTTAACAGACAGGGGTATAGGCCAGTGTCTC-T
 8210 8220 8230 8240 8250 8260 8270
 2190 2200 2210 2220 2230
 TTCAGACCCACCT-----CCCACCCCCGAGGGGACCCGAC-AGGCC--C---GAAGGAATAGAA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TCCCCCACCCCTTATTTCAGACCCATACCCAACAGGATCCGGCTTGCCAACCAAAGAAGGCAAAAAAA
 8280 8290 8300 8310 8320 8330 8340
 2240 2250 2260 2270 2280 2290 2300
 GAAGAAGGUTGGAGAGAGAGACAGAGACAGATC---CATTGATTAGTGAACGGATCCTTAGCACTTATCTG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GGAGACGGTGGAGGCAAGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCAATTTC-CTGATCCG
 8350 8360 8370 8380 8390 8400 8410
 2310 2320 2330 2340 2350 2360 2370
 GGACGATCTGGAGAGC---CTTG---TGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTA-A
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ---CCAACGTGATAAGGCTCTTGACCTGGCTATTCAAGCAA---CTGCAGAACCTTGCTATCGAGAGCATA
 8420 8430 8440 8450 8460 8470 8480
 2380 2390 2400 2410 2420 2430 2440
 CGAGGATTGGAACCTGCTGGACGCACGGGTGGGAAGCCCT-CAAATATTGGTGGAAATCTCCTACAGTAT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CCA-GATGCTCCACCAATAATCCAGAGGCTCTGCGACCCTACGGAGAATTGAGAA-GTCCT-CAGGCT
 8490 8500 8510 8520 8530 8540 8550
 2450 X
 TGGAGTCA-GGAACTAAG
 ||| | | | | | |
 TGAACGTACCTAACCAA
 8560 8570

8. KUNZ-158-FL3D, S60

HIV2R0DX Human immunodeficiency virus type 2 ROD isolate RN

ID HIV2R0DX standard; RNA; 5671 BP.
 XX
 AC X05291
 XX
 DT 04-JUN-1997 (annotation)
 XX
 DE Human immunodeficiency virus type 2 ROD isolate RNA genome
 DE (HIV-2)
 XX
 KW acquired immune deficiency syndrome; art gene; env gene; f gene;
 KW gag gene; pol gene; q gene; r gene; tat gene.
 XX
 OS Human immunodeficiency virus type 2
 OS ROD isolate
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-5671)
 RA Alison M.
 RT
 RL Submitted (03-JUN-1997) on tape to the EMBL Data Library by:
 RL Marc Alison, Unité d'Ecologie Virologie, and CNRS UMR1157, Institut

GCGTGGAAATGCGGACACATCTCCCTTGUGGATATTG-----ATG-ATCTGTAGTGCTACAGAAAAATTG
 TATT---AGCTAAGTCG---TTG---CTTAGTATATTGCACCCAATATGTAAGTGT-TTTCTATGGCGTACCCA
 6180 6190 6200 6210 6220 6230

130 140 150 160 170 180
 TG-GCTCAAGCTTATTATGGGGTACCTGTGGAAAGGAAGCAACCACCACTCTATTTGTG---CAT-CAG
 CGTGGAAAATGCAACCATTCGCCCTCTTTGTGCAA-CCAGAAATAGGGA---TACTTGGGGAACCATACAG
 6240 6250 6260 6270 6280 6290 6300

190 200 210 220 230 240 250
 ---ATGCTTAAGCAGATGATACAGAGGGTACATAATGTTTG--GGCCACACATGCCTGTGTACCCACAGACG
 TGCTTGCGCTGAAATGATTATCAGG-AAATAA-CTTGAAATGTAACAGAGGGCTTTG-ATGCATGGAAT
 6310 6320 6330 6340 6350 6360 6370

260 270 280 290 300 310 320
 CCAC-CGACAAAGAGATAGTATTCGTAATG---TGACAGAAAATTTAACATGTGGAAAAATGACATG-GTA
 AAATACAGTAACGAAACAGCAATAGAAGATGTCTGGCATCTATTGAGACAT---CAATAAAACCATGTGTC
 6380 6390 6400 6410 6420 6430 6440

330 340 350 360 370 380 390
 GAACAGATGCGTGGGGATATAATCAGTTATGGGATCAAAGCCTAAAGC---CATGTGTAAAATTAACCCCCAC
 AAACATGAAACACCTTTATGTTG-AGCAATGAAATGCAGCAGCACAGAGAGCAGCACAGGGAA---CAACACAAC
 6450 6460 6470 6480 6490 6500 6510

400 410 420 430 440 450 460
 TCTCTGTTTAAAGTGCGACTGATTGGGGAAATGCTACTAATACCAATA-CTAG---TAA-TACCAATAG
 -CTGAAAGAGGACAAA---GCACAAACCAACACACACCCAC---AGACCCAGGAGCAAGAGATAAGTGAGGATAC
 6520 6530 6540 6550 6560 6570

470 480 490 500 510 520
 TAGTAGCG---CGCAATGAA---TGATGGAGAAAGGAGAGATAAAAAAACTGCTCTTCAATATCAGCACAAGN
 TCCATGCCGACCGCGAGACAACTGCT-CAGGATTGGGAGAGGAAGAAACGATC---AATTGCCAGTTCAA-T
 6580 6590 6600 6610 6620 6630 6640

530 540 550 560 570 580 590
 ATAGGAGG-TAAAGTGGAGA-AAGAAATATGCATTTTTATAAACTTGATATAATACCAATAGATAATGATA
 ATGACAGGATTAAAGAGATAGAAAAAAACAGTATAAT-GAAACATGGTA---CTCAAAAGATGTGGTT
 6650 6660 6670 6680 6690 6700 6710

600 610 620 630 640 650 660
 CTACUCAGCCTATCGTTGACAAGTCTGAAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTGAGC
 GTGAGACATATAAGGACAAA---ATCAGACCCAGT-GTTACATGAACCATTCG-AACACATC---AGT
 6720 6730 6740 6750 6760 6770

670 680 690 700 710 720 730 740
 CAATTCGGATACATTATTGTRCCCGUGECTGGTTTGCATCTAAATGTAATAATA-AGACGTTCAATGGA
 CA---TCACAGAA-ATCA-TGTGACAAG---CACTATTGGGATGCTATAAGGTTAGATACTGTGCACCCACCGGG
 6780 6790 6800 6810 6820 6830

750 760 770 780 790 800
 ACAGAACCGATTCAGA---ATGTCAGCACAGTA---CA---ATGTACACA---TGGAATTAGGCCAGTAGTA
 TTATACCCCTTGTAGGATGTAATGAAACCA-ATTATTCAAGGCTTGCACCCAACCTGTTCTAAAG-TAGTAGCT
 6840 6850 6860 6870 6880 6890 6900

TCAAC-TCAAC---TCTTG-TGAACTGGCAAGTCAGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTCAC
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TCTACATGCACTGGATGATEGHARCCCAAACCTTCACATG---GTTGGCTTA-A--TGGCACTAGAGC
 6810 6820 6830 6840 6950 6960 6970
 870 880 890 900 910 920 930
 AGACAATGCTAAACAC---TAACTACTAAGCTGAACCAATCTGTAGAAATTAAATTGTACAAGACCCAACAAAC
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AGAGAAATACAAACATATATCTATTGCGATACGAGATAA---TAGAACT--ATCAT-CA-GCTTAAACA--
 9180 9190 9200 9210 9220 9230
 940 950 960 970 980 990 1000
 AATAACAAG/MA---AAAGATATCGTATCCAGAGGGGACCAGGGG---GAGCATTGTTAC-AATAGGAAAAAT--
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AATATTA1AATTCAGTTGCGATTGTAAGAGG ---CCAGGGAAATAAGACAGTGAAACAAATAATGCTTATGT
 7040 7050 7060 7070 7080 7090 7100
 1010 1020 1030 1040 1050 1060 1070
 -AGUAGATATGAGACAGTCAGATTAAACATTAGTAGAGCAAAATGCAATGCCACTTAAAACA---GATA
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CAGACGATGTTGACTCCACT---AC--CAGCGATCAATA---AAAGACCCAGACAAGCATGGTGCTG
 7110 7120 7130 7140 7150 7160
 1080 1090 1100 1110 1120 1130
 G---GTA---AGCAATTAGATA-ACAAT-TTGGAAAT-AATAAAACAATAATCTTAAGCAATCCTCAGG---
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GTTCAAAAGGCGAAATGCGAAAGACCGATGAGGAGGTGAAGGAAACCT---TGCAAAACATCC-CAGGTAT
 7170 7180 7190 7200 7210 7220 7230
 1140 1150 1160 1170 1180 1190
 AGGGACCCAGAAATTTAACGGACAGTTTAATTGTTGAG---GGGAATTTTC-TAC---TGTA---
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
 ABAGGAAAGU---AAT-1AAGAUAAACAAATTAGCTTGCAGCGCCAGGAAAAGGCTCAGACCCAGAAGTAGC
 7240 7250 7260 7270 7280 7290 7300
 1200 1210 1220 1230 1240 1250
 ATTCA---ACAC-GACTGTAACTACTTGGTTAAATAGTACTTGGAGTACTGAAGGGTCAAAT-AACAC
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
 ATAGATGTCAGTACATGCGAGG---AGTTTCTACTGCAACAT-GACT---TGGTTCTCAATTGGATAG
 7310 7320 7330 7340 7350 7360 7370
 1260 1270 1280 1290 1300 1310 1320
 TGAAAGAAAGTACA---CAAGACACTUCCATGCGAAATAAAACAATTATAAACATGTGGCAGGAAGTAGG
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
 AGAA-TAAGACAGCGCAATTATGC-ACCGTGCATATAAGCAAATAATTACACATGGCATAAGGTAGG
 7380 7390 7400 7410 7420 7430 7440
 1330 1340 1350 1360 1370 1380 1390
 AAAAGCAATG---TATGCCCCGCCATCAGCGGGACAAATTAGATGTTCATCAAATATTACAG---GGCTGC-TA
 ||||| ||| ||| ||| ||| ||| ||| |||
 ---GAGAAATGTAATTGCTCGCA---GGGA-AGGGGAGCTGTCCTGCAACT-CAACAGTAACCAGCATA
 7450 7460 7470 7480 7490 7500
 1400 1410 1420 1430 1440 1450
 TTAAACAAGAGATG---TGGTAATAACAAACAAATGGGTCCGAGAT---CTTCAGACCTGGAGGGAGG-AGA---TA
 ||||| ||| ||| ||| ||| ||| ||| |||
 ATTCGTA-ACATTCAGTGGCGAA-AACAATAATCAGACAAACATTACCTTAGTGC-AGAGGTGGCAGAACTA
 7510 7520 7530 7540 7550 7560 7570
 1460 1470 1480 1490 1500 1510 1520
 TGAGGAGACA-AT AGGAGAAAGTGTATTATATAAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCC
 ||||| ||| ||| ||| ||| ||| |||
 T---ACAGATTGAGATTGAGAGAT-TATAAAT---TGGTAGAAATAACACCAATTGGCTTCGCACT
 7580 7590 7600 7610 7620 7630
 1530 1540 1550 1560 1570 1580 1590

ACGTACGGCAAAAGA-AGAGTGTGAGAGAGAAAAAGAGCAGTGG-GAATAGGAGCTTGTTCCTTGGGT
 ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AGCTTAAAGA-CGAGATCTCTCTGTCACGGGAGACATAAACAGAGGTGTGTTCGTGCTAGGGTTCTTGGGT
 7640 7650 7660 7670 7680 7690 7700

 1600 1610 1620 1630 1640 1650 1660 1670
 TCT-TGGGAGCAGCAAGAAGAAGTATGGGCAACGGTCAATGACCGCTGACGGTACAGGCCAGACAATTATTG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TTTCTGGCAACACCAEGTTCTGCAATGGGCGGGCGTCCCTGACCGTGTGTCGGCTCAGTCCCAGACTTTACTG
 7710 7720 7730 7740 7750 7760 7770

 1680 1690 1700 1710 1720 1730 1740
 TCTGTTATAUTLCAGGAGCAGAACGATTTGCTGAGGGCTATTGAGGCAGCAACAGCATCTGTTGCAACTCACA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GCGCGGATAGCTGAGCAACAGCAACAGCTGTTGGACGGTCAAGAGACAAACAAGAACTGTTGCGACTGACC
 7780 7790 7800 7810 7820 7830 7840 7850

 1750 1760 1770 1780 1790 1800 1810
 GTCTGGGGGATCAGGCGCTCCAGGCAAGAATCCTGGCTGTGGAAAAGATACTAAAGGATCAACAGCTCCTG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GTCAGGGGAACTCCAGGCAAGACTGCTATAGAGAAGTACCTACAGGACCAGGCGCGGGCT-
 7860 7870 7880 7890 7900 7910 7920

 1820 1830 1840 1850 1860 1870 1880
 CGGATT--TGGGGTTGCTCTGGAAAACCTCATTTGCAACCACTGCTGTGCCCTTGGAAATGCTAGTTGGAGTAATA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 -AAATTCACTGGGATETGCGTTTAGACAAGTCTGCCACACTACTGTACCATGG-----GT-----TAATG
 7930 7940 7950 7960 7970 7980

 1890 1900 1910 1920 1930 1940 1950
 ATATCTGGAAAGATTTGGAAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACAATTACACAAGCT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ATTCCTTACACCTGACTGGAAACATATGACGTGGCAGGAATGGGAAAAACAAAGT--CCGCTACCTGGAGGC
 7990 8000 8010 8020 8030 8040 8050

 1960 1970 1980 1990 2000 2010 2020
 TAATA-CACTGGTTAATT-GAAGAATCAGAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAAATTAGA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AAATATCACTGGAACTTAAAGAACGGCACAAATTCAAGCAAGAGAAAAATATGTATGAACATACAAAAATTAAA
 8060 8070 8080 8090 8100 8110 8120

 2030 2040 2050 2060 2070 2080 2090
 TAATTGGGGGAAAGTTGGAAATTGGTTAACATAACAAATTGGCTGTGGTATA-TAAAAAT--ATTCTATAAT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TAGGGGAACTTATTTGGCAATTGGTTGACTTAACCTCCTGGGTCAAGTATATTCAATATGGAGTGCTTAT
 8130 8140 8150 8160 8170 8180 8190

 2100 2110 2120 2130 2140 2150 2160
 GATAGTGGGAAAGTTGGAAATTGGTTAACAGATAGTTTGCTGTACTTCTATAGTGAATAGAGTTAGGCAGGG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AATAGTGGGAAAGTTAGAGCTTAAGAATAGTGTATATGTAGTACAAATGTTAAGTAGGCTTAGAAAGGG
 8200 8210 8220 8230 8240 8250 8260

 2170 2180 2190 2200 2210 2220
 ATATTCACG-ATTATCGTTTC-----AGACCCACCC---TCCCAACCCCCGAGGGGACCC--CGACAG-
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CTATGGGCTTCTTCTTCCCGGGTTATATCCAACAGATCCATATCCACA-AGGACCGGGGACAGC
 8270 8280 8290 8300 8310 8320 8330

 2230 2240 2250 2260 2270 2280
 ---GGG---CGGGGAAATAGAGAACGGGAGGGATGGAGAGAGAGACAGATCCATTGCGATTAGTGAACGGAT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CGGGGAAACGGGAAAGAACGGGAGGGAGGGATGGAGAGAGACAGATACTGGCCCTGGCCGATAGCAT
 8340 8350 8360 8370 8380 8390 8400

 2290 2300 2310 2320 2330 2340 2350

9. KUNZ-158-CLXG SET

RESIV251 Human Immunodeficiency virus (Mac251) envelope gene

ID PRHSIV251 standard: DNA; 1142 BP.
 XX
 AC M00379; Y00294;
 XX
 DT 23-JUN-1988 (acc# added)
 DT 27-MAY-1992 (annotation)
 XX
 DE Simian immunodeficiency virus (Mac251) envelope gene DNA (part.)
 DE integrated copy
 XX
 KW env gene; envelope gene.
 XX
 OS Simian immunodeficiency virus
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN 111 (bases 1-1142)
 RA Kestler H. W.;
 RT
 RL Submitted (25-FEB-1988) to the EMBL Data Library by:
 RL Kestler H. W., Harvard Medical School, New England Regional Primate
 Research Center, Department of Microbiology, One Pine Hill Drive,
 RL Southborough, Mass. 01772, USA.
 XX
 RN 121
 RA Kestler H. W., Li Y., Naidu Y. M., Butler C. V., Ochs M. F.,
 RA Javinek P., King N. V., Daniel M. D., Desrosiers R. C.;
 RT "Comparison of Simian immunodeficiency virus isolates";
 RL Nature 331:619-622(1988).
 XX
 CC #source= strain=Macaca mulatta 251(host); clone=lambda SIV 251;
 XX
 FH Key From To Description
 FH
 FT SITE 1 1142 put. env gene
 (1 is 2nd base in codon)
 FT SITE 1026 1028 in frame stop codon
 XX
 SQ Sequence 1142 BP: 322 A; 208 C; 273 G; 231 T; 108 other;

Initial Score = 259 Optimized Score = 565 Significance = 0.00
 Residue Identity = 51% Matches = 622 Mismatches = 488
 Gaps = 83 Conservative Substitutions = 0

1180 1170 1180 1190 1200 1210 1220

1 1111 1111 1111 11111111 11 11 11 11 11 11 11 11
 TGTGAGACAAATTGAGAGGAGAATTCTCTACTGTAA---AA-TGAATTGGTTCTA-AATTGGGTAGAGGA
 X 10 20 30 40 50 60
 1230 1240 1250 1260 1270 1280 1290
 TGAATCTTGGAACTACTGAABGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAA
 123 124 125 126 127 128 129
 TAAAGATGTTACGTTACCCAGAAGCCAAAGGA-AC---GCCATAGAAGGAATTAC-GTGCCGTGTCATATTAG
 75 80 85 100 110 120 130
 1300 1310 1320 1330 1340 1350 1360
 ACATTTATAAACATGGCAGGAAGTAGGAAAGCAATG--TATGCCCTCCCATCAGCGGACAAATTAGA
 130 131 132 133 134 135 136
 ACACATAATCAGAACACTTGGCATAAAAGTAGGCAAA--AATGTTATTCGCTCCAAGAGAGGGAGACCTCACG
 140 150 160 170 180 190 200
 1370 1380 1390 1400 1410 1420 1430
 TTTTCACTAAATTACAGGGCTCTTAAACAAGAGATEGTTAATAACAACAATGGGTCCGAGATCTTC
 137 138 139 140 141 142 143
 TGTAACTCAGTGGACCGATCTCATAGCAAACATAAGATTGGACTGATGGAAACCA---AACTAATATCACC
 210 220 230 240 250 260 270
 1440 1450 1460 1470 1480 1490 1500
 AGACUTG---GAUAGGG-AGATATGGGACAATTGGAGAAGTGAATTATAAAATATAAAAGTAGTAAAAATT
 144 145 146 147 148 149 150
 ATGAGTGCAGATGGCAGACT---CTATCGATTGGAGTTGGGAGAT-TATAAT-----TAGTAGAGATN
 280 290 300 310 320 330
 1510 1520 1530 1540 1550 1560 1570
 GAACCATTTGGAGTACACCCACCAAGGCAAAAGAG---AAGAGTGGT-GCA--GAGAGAAAAAGAGCAGT
 151 152 153 154 155 156 157
 NNN
 340 350 360 370 380 390 400
 1580 1590 1600 1610 1620 1630 1640
 GGAAATAAGAGCTTTCTTGGGTTCT-TGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGC
 158 159 160 161 162 163 164
 GGGCTTTGGCTAACGGTTCTGGGTTCTGGCAATGGGCGGGGTGTCATGGGCGGGTGNNNNNNNNN
 410 420 430 440 450 460 470
 1650 1660 1670 1680 1690 1700 1710
 TGGGTTACAGGCCCCAAATTATGTCTGGTATAGTCAGCAGCAGAACAAATTGCTGAGGGCTATTGAGG
 165 166 167 168 169 170 171
 NNNACGCTCAATCCCCAACCTTATGGCTGGGATAGTCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGA
 480 490 500 510 520 530 540
 1720 1730 1740 1750 1760 1770 1780
 CGCAACAGUATCTGTTGCAAUTCAACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAACCTGGCTGTGGAAA
 172 173 174 175 176 177 178
 AACACAAAGAAATTGTTGCAACTGACCGCTCTGGGGAAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGA
 550 560 570 580 590 600 610 620
 1790 1800 1810 1820 1830 1840 1850 1860
 GATACCTAAAGGATCAACAGCTCTGGGGATTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGC
 179 180 181 182 183 184 185 186
 AGTACTTAAAGGACGCGCAGACTGAATGCTGGGGATGTGCGTTAGACAAGTCTGCCACACTACTGTAC
 630 640 650 660 670 680 690
 1870 1880 1890 1900 1910 1920
 CTTAATG---ATTAATGTTGGAGTAATAAAATCTCTGGAAACAGATTGGAAATAACATGACCTGGATGGAGTGGG
 187 188 189 190 191 192
 CAACGCGATATGAGT---CTAACCCAGACTGGAAACA-AT---GA-TA-----CTTGGCAAGAGAGTGGG
 700 710 720 730 740 750
 1930 1940 1950 1960 1970 1980 1990 2000
 AAAGGCGGGCTTCTGAGGAAATGAGGAAATCGCAAAACAGCAAGGAAAG

ACGGAAAGGTTTCACTTCCTTGAGCAAAATATAACA-GCCCTCCTAGAAGAGGGACAAATTCAACAAGAGAAG
 780 770 780 780 800 810 820
 2010 2020 2030 2040 2050 2060 2070
 AATGAAACAAGAATTATGGAAATTAGATAAATGGCAAGTTGTGG-AATTGGTTAACATAACAAATTGGCT
 830 840 850 860 870 880 890
 2080 2090 2100 2110 2120 2130 2140
 GTCGTATACTAATATTCATATGATAGTAGGAGGCTTGGTAGGTTAACATAGTTTGCTGACTTC
 900 910 920 930 940 950 960
 2150 2160 2170 2180 2190 2200 2210
 TATGAGTGAATAGGTTT-AGGCAGGGATATTCAACCATATCGTTTCAAGACCCACCTCCAAACCCCGAGGGG--
 970 980 990 1000 1010 1020 1030
 AATGCTAATCTA-AGTTAAGGCAGGGATAGGCCAGTGT--TCTCTCCCCACCCCTCTTATTCTAGCAGAC
 1040 1050 1060 1070 1080 1090 1100
 2220 2230 2240 2250 2260 2270
 ---ACCCAGCAGG-UCCG---AAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTGAT
 1110 1120 1130 1140 X
 TCAATACCCAAACAGGACCCGGCACTGCCAACAG-AGAAGGCAAAGAAGGAGACGGTG-GAGA---AGGCGGT
 2280 2290 2300 2310 2320
 TAGTGAACGATCCTTAGCACTTATCTG--GGACGATCTGCGGAGC
 1150 1160 1170 1180 1190
 ---GTAACGAGGCG-TGCG-CTTCGAGATAGAATATATTCAATTTC

10. KUNZ-158-GL3.2_SE18

M15127 Figure 1. Structure of the art gene of HTLV-III, C

ID M15127 unannotated; xxxx; 306 BP.
 XX
 AC M15127
 XX
 DT 10-JUL-1993 (Incorporated)
 XX
 DE Figure 1. Structure of the art gene of HTLV-III, Coding Exon II.
 XX
 KW -
 XX
 OS
 OC
 XX
 RN (1) (bases 1-306)
 RA Bob W.C., Sojka J.G., Rosen C.A., Haseltine W.A.;
 RT "Expression of the art Gene Protein of Human T-Lymphotropic Virus
 Type III (HTLV-III/LAV) in Bacteria";
 RL J. Virol. 61:633-637(1987).
 XX
 FH Key From To Description
 FH
 XX
 SQ Sequence 306 BP: 65 A; 63 C; 87 G; 66 T; 0 other;

Initial Score = 184 Optimized Score = 298 Significance = 0.00
 Residue Identity = 99% Matches = 299 Mismatches = 2
 Gaps = ! Conservative Substitutions = 0

Kunz
07/158452
Claim 32
Genbank

Results file Kunz-158-CL32nh.res made by sheppard on Thu 8 Mar 90 11:09:14-PST.

Query sequence being compared: KUNZ-158-CL32. SEQ
Number of sequences searched: 31228
Number of scores above cutoff: 38

Results of the initial comparison of KUNZ-158-CL32. SEQ with:
Data bank : GenBank 62.0; all entries

100000+
N
U50000+
M
B
E
R
→*
O
F10000+
S
E 5000+
Q
U
E
N
C
E
S 1000+
→* w
500+
→*
→*

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230
TGTGCGAAGCTATTCACCATTTATCCTTCAAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGG
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TAGCGAGGTTATTCACCATTTATCCTTCAAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGG
X 10 20 30 40 50 60 70

2240 2250 2260 2270 2280 2290 2300
AATAGAAAGAGAGAGAGAGAGAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AATAGAAAGAGAGAGAGAGAGAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
80 90 100 110 120 130 140

2310 2320 2330 2340 2350 2360 2370
CTGGGACGCTTGCGGAGGCCCTTGCGCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CTGGGACGATCTGGGAGCC-TGTGCGCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
150 160 170 180 190 200 210

2380 2390 2400 2410 2420 2430 2440
GGATTGTGAAACTTCTGGGACCGACGGGGTGGGAAGGCCCTCAAATATTGGTGGAATCTCCTACAGTATTGGA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GGATTGTGAAACTTCTGGGACCGACGGGGTGGGAAGGCCCTCAAATATTGGTGGAATCTCCTACAGTATTGGA
240 250 260 270 280

2450 X
CTGGGAACTTAAAG
||||| |||||
CTGAGGAGATTAAGG
290 300

A scatter plot showing the relationship between Score (X-axis) and STDEV (Y-axis). The X-axis ranges from 0 to 2341 with major ticks every 200 units. The Y-axis ranges from 0 to 50 with major ticks every 10 units. There are 11 data points plotted, all marked with an asterisk (*). The points show a positive correlation, with STDEV increasing as Score increases.

Score	STDEV
111	1
111	1
111	1
111	1
111	1
111	1
111	1
111	1
111	1
111	1
111	1
520	1
760	1
1040	1
1301	1
1561	1
1821	1
2081	1
2341	1

PARAMETERS

Similarity metric	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	147		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean 37	Median 37	Standard Deviation 19.85
Times:	CPU 00:32:54.08	Total Elapsed 02:09:12.00	
Number of realizations:	37183850		

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

2.	HIVPV22	***** 22 standard deviations above mean ****	Human immunodeficiency virus t	9770	1877	2190	92.71	0
3.	HIVHXB3	Human immunodeficiency virus t	3156	1873	2186	92.51	0	
4.	HIVBH102	Human immunodeficiency virus t	8932	1872	2186	92.46	0	
		***** 51 standard deviations above mean ****						
5.	HIVHXB2CG	Human immunodeficiency virus t	9718	1858	2174	91.75	0	
		***** 55 standard deviations above mean ****						
6.	HIVNL43	Human immunodeficiency virus t	9709	1729	2181	85.25	0	
		***** 60 standard deviations above mean ****						
7.	HIVELT00	Human immunodeficiency virus t	9176	1246	1904	60.91	0	
		***** 66 standard deviations above mean ****						
8.	HIVSC	Human immunodeficiency virus t	4273	1159	2151	56.53	0	
		***** 64 standard deviations above mean ****						
9.	HIVH3BHR8	Human immunodeficiency virus t	3563	1112	1773	54.16	0	
		***** 51 standard deviations above mean ****						
10.	HIVZ321	Human immunodeficiency virus t	3457	1066	1970	51.85	0	
		***** 48 standard deviations above mean ****						
11.	HIVMN01	Human immunodeficiency virus t	9738	996	2223	48.32	0	
		***** 47 standard deviations above mean ****						
12.	HIVZ22G	Human immunodeficiency virus t	9081	983	1920	47.66	0	
13.	HIVZ6	Human immunodeficiency virus t	5159	983	1924	47.66	0	
14.	HIVTH32	Human immunodeficiency virus t	2903	975	1385	47.26	0	
15.	HIVJY1	Human immunodeficiency virus t	2653	972	1936	47.11	0	
		***** 45 standard deviations above mean ****						
16.	HIVCDC02	Human immunodeficiency virus t	3373	932	2209	45.09	0	
		***** 44 standard deviations above mean ****						
17.	HIVMAL	Human immunodeficiency virus t	9229	916	2060	44.29	0	
		***** 42 standard deviations above mean ****						
18.	HIVPFERW	Human immunodeficiency virus ty	2622	887	1587	42.83	0	
19.	HIVRF	Human immunodeficiency virus t	9128	887	1587	42.83	0	
20.	HIVSF200	Human immunodeficiency virus t	9737	883	1960	42.63	0	

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init.	Opt.	Sig.	Frame
			Score	Score		
1. HIVERU1	Human immunodeficiency virus t	9229	2341	2456	0.00	0
2. HIVMN01	Human immunodeficiency virus t	9738	996	2223	0.00	0
3. HIVCDC02	Human immunodeficiency virus t	3373	932	2209	0.00	0
4. HIVPV22	Human immunodeficiency virus t	9770	1877	2190	0.00	0
5. HIVBH102	Human immunodeficiency virus t	8932	1872	2186	0.00	0
6. HIVHXB3	Human immunodeficiency virus t	3156	1873	2186	0.00	0
7. HIVNL43	Human immunodeficiency virus t	9709	1729	2181	0.00	0
8. HIVHXB2CG	Human immunodeficiency virus t	9718	1858	2174	0.00	0
9. HIVSC	Human immunodeficiency virus t	4273	1159	2151	0.00	0
10. HIVMAL	Human immunodeficiency virus t	9229	916	2060	0.00	0
11. HIVZ321	Human immunodeficiency virus t	3457	1066	1970	0.00	0
12. HIVSF200	Human immunodeficiency virus t	9737	883	1960	0.00	0
13. HIVJY1	Human immunodeficiency virus t	2653	972	1936	0.00	0
14. HIVZ6	Human immunodeficiency virus t	5159	983	1924	0.00	0
15. HIVZ22G	Human immunodeficiency virus t	9081	983	1920	0.00	0
16. HIVELT00	Human immunodeficiency virus t	9176	1246	1904	0.00	0
17. HIVH3BHR8	Human immunodeficiency virus t	3563	1112	1773	0.00	0
18. HIVPFERW	Human immunodeficiency virus ty	2622	887	1587	0.00	0
19. HIVRF	Human immunodeficiency virus t	9128	887	1587	0.00	0
20. HIVSERV1	Human immunodeficiency virus t	3600	869	1479	0.00	0

HIVREF002	Human immunodeficiency virus type 1, isolate BRU,			
LOCUS	HLVREF002	9229 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate BRU, complete genome (LAV-1).			
ACCESSION	K02813			
KEYWORDS	TAT protein; TAT protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activation.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate BRU (LAV-1), proviral DNA clone Lambda-J18.			
ORGANISM	Human immunodeficiency virus type 1 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	1 (bases 1 to 9229)			
AUTHORS	Wetmore, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M.			
TITLE	Nucleotide sequence of the AIDS Virus, LAV			
JOURNAL	Cell 40: 9-17 (1985)			
STANDARD	full staff_review			
REFERENCE	2 (bases 1712 to 1749) revision of [1]			
AUTHORS	Alizon, M., Watt, Hobson, S., Montagnier, L. and Sonigo, P.			
TITLE	Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients			
JOURNAL	Cell 48: 63-74 (1986)			
STANDARD	full staff_review			
COMMENT	<p>The original LAV, sometimes called LAV-1 to distinguish it from HIV-2 (LAV-2), is now referred to as HIV-1bru. An infectious clone of this virus has been constructed by Keith Peden, Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205 (301) 955-3652. HIVNL43 is also an infectious clone having for its 3' half a clone of the BRU isolate.</p> <p>Acquired immune deficiency syndrome (AIDS) is caused by a retrovirus known by several different names, probably representing two separate strains: human T-cell lymphotropic virus-III (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the other. All three viruses, whose sequences do not differ by more than about 8%, are believed to belong to the retroviral subfamily Lentiviridae, or "slow" viruses.</p> <p>For the details of the annotation and for other pertinent references, see the HIV reference entry.</p>			
FEATURES	from	to/span	description	
pept	336	1874	gag polyprotein	
pept	< 1631	4678	pol polyprotein (NH2-terminus uncertain; AA at 1631)	
pept	4823	5201	vif protein	
pept	5141	5431	vpr protein	
pept	5412	5626	tat protein, exon 2 (first expressed exon)	
pept	7572	8017	tat protein, exon 3 (AA at 7973)	
pept	6551	5626	rev protein, exon 2 (first expressed exon)	
pept	7872	8246	rev protein, exon 3 (AA at 7974)	
pept	5643	5888	vpu protein	
pept	5803	6383	envelope polyprotein	
pept	6350	6010	nef protein	
pre-msg	1	8223	genomic mRNA	
pre-msg	1	9228	tat, rev, nef subgenomic mRNA	
IVS	250	5358	tat, rev, nef subgenomic mRNA intron 1	
IVS	5627	7871	tat cds intron 2	
IVS	5627	7871	rev cds intron 2	
IVS	5627	7871	tat, rev, nef subgenomic mRNA intron 2	
LTR	< 1	130	5' LTR	
LTR	3673	> 9229	3' LTR	
rpt	< 1	97	R repeat 5' copy	
rpt	9133	9229	R repeat 3' copy	
binding	192	198	primer (Lys-tRNA) binding site	

revision 1712 1749 atttcttcagagcagaccagccaaacagccccaccag in [2];
signal 9205 9210 mRNA polyadenylation signal
BASE COUNT 3263 a 1656 c 2232 g 2052 t
ORIGIN Cap site of genomic RNA.

Initial Score = 2341 Optimized Score = 2456 Significance = 0.00
Residue Identity = 99% Matches = 2456 Mismatches = 3
Gaps = 1 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
AAGAGCAGAAGACAGTGGCAATGAGAGTGAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG							
AAGAGCAGAAGACAGTGGCAATGAGAGTGAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG							
X	5790	5800	5810	5820	5830	5840	5850

```

          80      90      100     110     120     130     140
GCACCATGCTCCTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  

GCACCATGCTCCTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG
5860      5870      5880      5890      5900      5910      5920

```

150	160	170	180	190	200	210
TACCTGTGTGGAAGGAAGCAACCACCTCTATTTGTGCATCAGATGCTAAAGCATATGATA						
TACCTGTGTGGAAGGAAGCAACCACCTCTATTTGTGCATCAGATGCTAAAGCATATGATA						
5930	5940	5950	5960	5970	5980	5990

220	230	240	250	260	270	280	
ATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA							
6000	6010	6020	6030	6040	6050	6060	6070

290	300	310	320	330	340	350	360
CAGAAAATTTAACATGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTATGGGATC							
CAGAAAATTTAACATGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTATGGGATC							
6080	5090	6100	6110	6120	6130	6140	

3'70	380	390	400	410	420	430
AAAGCCTAAGCCATGTTGTAATAACCCCCACTCTGTGTTAGTTAAAGTGCACTGATTGGGAATGCTA						
AAAGCCTAAGCCATGTTGTAATAACCCCCACTCTGTGTTAGTTAAAGTGCACTGATTGGGAATGCTA						
6150	6160	6170	6180	6190	6200	6210

440	450	460	470	480	490	500
CTAATACCAATTACTGAAATACCAATAGTAGTACGGGGAAATGATGATGGAGAAAGGAGAGATAAAAAACT						
.....						
CTAATACCAATTACTGAAATACCAATAGTAGTACGGGGAAATGATGATGGAGAAAGGAGAGATAAAAAACT						
6220	6230	6240	6250	6260	6270	6280

510 520 530 540 550 560 570
 GCTCTTTCYATATCAGCACACAGATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTATAAACTTGATA

 GCTCTTTCGATATCGACACACAGCAAAAGGTAAAGGTGCAGAAAGAATATGCATTTTTTATAAACTTGATA
 6290 6300 6310 6320 6330 6340 6350

580 590 600 610 620 630 640
 TAATACCAATTATAAATATACTAUCAGCTATACETTGACAAGTTGTAACACACCTCAGTCATTACACAGGCC
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
 TAATACCAATTATAAATATACTAUCAGCTATACETTGACAAGTTGTAACACACCTCAGTCATTACACAGGCC
 6360 6370 6380 6390 6400 6410 6420 6430

GTGCTTAAAGATTAATCTTGACGCCATTCCCATACATTATTGTCCCCGGCTGGTTTGCGATTCTAAATGTA
6440 6450 6460 6470 6480 6490 6500
730 740 750 760 770 780 790
ATAATAAGACGGTCAAGACACGACGACATGTCAGCACAGTACAATGTACACATGGAATTAGGC
ATAATRAACGCGTTCAAACGACGACGACATGTCAGCACAGTACAATGTACACATGGAATTAGGC
6510 6520 6530 6540 6550 6560 6570
800 810 820 830 840 850 860
CAGTAGATACTGACTCTCTTGTGAGTGGCTAGCTAGCAGAAGAAGAGGGTAGTAATTAGATCTGCCAATT
CAGTAGATACTGACTCTCTTGTGAGTGGCTAGCAGAAGAAGAGGGTAGTAATTAGATCTGCCAATT
6580 6590 6600 6610 6620 6630 6640
870 880 890 900 910 920 930
TCACAGACAAATCTAAACACATAATAGTACAGCTGACCAATCTGTAGAAATTAAATTGTACAAGACCCAACA
TCACAGACAAATCTAAACACATAATAGTACAGCTGACCAATCTGTAGAAATTAAATTGTACAAGACCCAACA
6650 6660 6670 6680 6690 6700 6710
640 650 660 670 680 690 700
ACAAATACAAATGAAACATATCGGATCCAGAGGGGACCGGGAGAGCATTGTTACAATAGGAAAAATAGGAA
ACAAATACAAATGAAACATATCGGATCCAGAGGGGACCGGGAGAGCATTGTTACAATAGGAAAAATAGGAA
6720 6730 6740 6750 6760 6770 6780 6790
1010 1020 1030 1040 1050 1060 1070 1080
ATATGAGAAAGGAAATTGTAACTTACAGACAAAATGCUATGCCACTTTAAACAGATAGCTAGCAAAT
ATATGAGAAAGGAAATTGTAACTTACAGACAAAATGCUATGCCACTTTAAACAGATAGCTAGCAAAT
6800 6810 6820 6830 6840 6850 6860
1090 1100 1110 1120 1130 1140 1150
TAATGAGAAAGGAAATTGTAACTTACAGACAAAATGCUATGCCACTTTAAACAGATAGCTAGCAAAT
TAATGAGAAAGGAAATTGTAACTTACAGACAAAATGCUATGCCACTTTAAACAGATAGCTAGCAAAT
6870 6880 6890 6900 6910 6920 6930
1160 1170 1180 1190 1200 1210 1220
CGCACAGTTTVAATTGAAAGGAAATTGTTCTACTGTAATTCAACACAACTGTTAACAGTACTTGGTTA
CGCACAGTTTVAATTGAAAGGAAATTGTTCTACTGTAATTCAACACAACTGTTAACAGTACTTGGTTA
6940 6950 6960 6970 6980 6990 7000
1230 1240 1250 1260 1270 1280 1290
ATATGACTTGTGAACTGAAAGGAAATTGAAAGGAAACTGACACAATCACACTCCCAGCAGAATAA
ATATGACTTGTGAACTGAAAGGAAATTGAAAGGAAACTGACACAATCACACTCCCAGCAGAATAA
7010 7020 7030 7040 7050 7060 7070
1330 1340 1350 1360
AAAGATTTTGTAACTTGTGAAAGGAAACCAATGTATGCCCTCCCATCAGCGGACAAATTAGAT
AAAGATTTTGTAACTTGTGAAAGGAAACCAATGTATGCCCTCCCATCAGCGGACAAATTAGAT
7080 7090 7100 7110 7120 7130 7140 7150
1370 1380 1390 1400 1410 1420 1430 1440
GTTCAACGAAATGAACTGAAAGGAACTGAAATTGAAAGGAAACTGACACAATGGGTCCGAGATCTTCA
GTTCAACGAAATGAAAGGAACTGAAATTGAAAGGAAACTGACACAATGGGTCCGAGATCTTCA
7160 7170 7180 7190 7200 7210 7220
1460 1470 1480 1490 1500 1510
GAACTGAAAGGAAACTGAAATTGAAAGGAAACTGACACAATGGGTCCGAGATCTTCA

1520 1530 1540 1550 1560 1570 1580
 CATTCTTCAATGAGAACCAAGGAAAGAGAATGGTCAGAGAGAAAAAGAGCAGTGGGAATAGGAG
 CATTTGCGTCTTGTGAGAACCAAGGAAAGAGAATGGTCAGAGAGAAAAAGAGCAGTGGGAATAGGAG
 7500 7510 7520 7530 7540 7550 7560
 1590 1600 1610 1620 1630 1640 1650
 CTTTTCTTCTTGTGAGAACCAAGGAAAGAGAATGGTCAGAGAGAAAAAGAGCAGTGGGAATAGGAG
 CTTTTCTTCTTGTGAGAACCAAGGAAAGAGAATGGTCAGAGAGAAAAAGAGCAGTGGGAATAGGAG
 7570 7580 7590 7600 7610 7620 7630
 1660 1670 1680 1690 1700 1710 1720
 CCALACAACTTCTTGCTGAGAACCAAGGAAACAATTGCTGAGGGCTATTGAGGCGAACAGCATC
 CCALACAACTTCTTGCTGAGAACCAAGGAAACAATTGCTGAGGGCTATTGAGGCGAACAGCATC
 7440 7450 7460 7470 7480 7490 7500
 1730 1740 1750 1760 1770 1780 1790 1800
 TGTTCAGCTTCAAGCTTGTGGCATCAACAGCTCCAGGCAAGAACATCCTGGCTGTGGAAAGATACTAAAGG
 TGTTCAGCTTCAAGCTTGTGGCATCAACAGCTCCAGGCAAGAACATCCTGGCTGTGGAAAGATACTAAAGG
 7540 7550 7560 7570 7580
 1810 1820 1830 1840 1850 1860 1870
 ATCAACGAGATCTTGGGATTCTGGGTTCTTGGAAAACCTATTTGCACCACTGCTGTGCCTTGGAAATGCTA
 ATCAACGAGATCTTGGGATTCTGGGTTCTTGGAAAACCTATTTGCACCACTGCTGTGCCTTGGAAATGCTA
 7590 7600 7610 7620 7630 7640 7650
 1880 1890 1900 1910 1920 1930 1940
 GTTGGACTTAATGATCTTGGGAAAGATTTGGAAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACA
 GTTGGACTTAATGATCTTGGGAAAGATTTGGAAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAAACA
 7660 7670 7680 7690 7700 7710 7720
 1950 1960 1970 1980 1990 2000 2010
 ATTACACAYABC1TAATACATTTCTTAAATGAAATAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAAATTAT
 ATTACACAYABC1TAATACATTTCTTAAATGAAATAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAAATTAT
 7730 7740 7750 7760 7770 7780 7790
 2020 2030 2040 2050 2060 2070 2080
 TGGAAATTAAATGATCTTGGGAAAGTTTGGAAATTTGCTGTGGTATATAAAATAT
 TGGAAATTAAATGATCTTGGGAAATTTGCTGTGGTATATAAAATAT
 7800 7810 7820 7830 7840 7850 7860 7870
 2090 2100 2110 2120 2130 2140 2150 2160
 TCATATATGATGAGGGCTTGGTAAAGAATAGTTTGTACTTCTATAGTGAATAGAGTTA
 TCATATATGATGAGGGCTTGGTAAAGAATAGTTTGTACTTCTATAGTGAATAGAGTTA
 7880 7890 7900 7910 7920 7930 7940
 2170 2180 2190 2200 2210 2220 2230
 AGGAGGGATTTCAACATTATGCTTCAACCCACCTCCAAACCCCCGAGGGGACCCGACAGGCCCGAAGGAA
 AGGAGGGATTTCAACATTATGCTTCAACCCACCTCCAAACCCCCGAGGGGACCCGACAGGCCCGAAGGAA
 7950 7960 7970 7980 7990 8000 8010
 2240 2250 2260 2270 2280 2290 2300

TATTCAGAATTCGCTTGTGAGACAGCAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCT
 8070 8080 8090 8100 8110 8120 8130 8140 8150
 2310 2320 2330 2340 2350 2360 2370
 GGGAGGATTCAGGAGCTTGTGAGACAGCAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCT
 8080 8100 8110 8120 8130 8140 8150
 2380 2390 2400 2410 2420 2430 2440
 ATTCTGGAACTCTGGAGACAGGGGTGGGAAGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGAGT
 ATTCTGGAACTCTGGAGACAGGGGTGGGAAGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGAGT
 8160 8170 8180 8190 8200 8210 8220
 2450 X
 CAGGAACTTAAAG
 CAGGAACTTAAAG
 8230 8240

2. KUNZ-158-CL32, SER

HIVMNCE Human immunodeficiency virus type 1, isolate MN, C

LOCUS HIVMNCE 9733 bp ss-RNA **VRL** 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate MN, complete genome.
ACCESSION M17448
KEYWORDS
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate MN, proviral DNA.
ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
REFERENCE 1 (bases 1 to 9733)
AUTHORS Bango,C., Guo,H.-G., Franchini,G., Aldovini,A., Collalti,E.,
 Ferrell,K., Wong-Staal,F., Gallo,R.C. and Reitz,M.S.Jr.
TITLE Envelope Sequences of two new United States HIV-1 isolates
JOURNAL Virology 164: 531-536 (1988)
STANDARD full staff_review
COMMENT Sequence kindly provided in computer readable form by Marv Reitz,
 N.C.I., Bethesda, MD. 20892 U.S.A.
 The MN isolate was taken from a pediatric AIDS patient in 1984.
 The pol coding sequence shows an in-frame stop codon at position 3783; the nef protein is prematurely truncated at position 9357.
 The vpu protein is prematurely truncated at position 6142.

FEATURES	from	to/stop	description
pept	787	2307	gag polyprotein
pept, pol	2091	5111	pol polyprotein (NH2-terminus uncertain; AA at 2091; in-frame stop codon at 3783)
pept	5056	5814	vif protein
pept	5574	5884	vpr protein
pept	5646	8059	tat protein, exon 2 (first expressed exon)
pept	8388	8486	tat protein, exon 3 (AA at 8397)
pept	8934	8059	rev protein, exon 2 (first expressed exon)
pept	8935	8607	rev protein, exon 3 (AA at 8398)
pept	8076	6144	vpu protein (premature termination)
pept	8238	5804	envelope polyprotein
pept	8811	9350	nef protein (premature termination at 9357 relative to other HIV-1 sequences)
pre-mseq	454	5855	genomic mRNA
pre-mseq	454	5655	tat, rev, nef subgenomic mRNA
IVS	740	5791	tat, rev, nef subgenomic mRNA intron 1
IVS	6090	8367	tat cds intron 2
IVS	6090	8367	rev cds intron 2

IV-3	6060	3393	tat, rev, nef subgenomic mRNA intron 2
LTR	1	633	5' LTR
LTR	3106	8730	3' LTR
rpt	453	550	R repeat 5' copy
rpt	2558	5515	R repeat 3' copy
binding	376	365	Spl binding site III
binding	387	396	Spl binding site II
binding	388	407	Spl binding site I
binding	635	652	primer (Lys-tRNA) binding site
site	3783	3785	pol cds in-frame stop codon
signal	8631	8636	mRNA polyadenylation signal
BASE COUNT	3463 a	1769 c	2344 g 2162 t

ORIGIN Left end of viral genome

Initial Score = 598 Optimized Score = 2223 Significance = 0.00
 Residue Identity = 90% Matches = 2253 Mismatches = 170
 Gaps = 58 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
AACAGCAAAACACGTTGCAATTAAGTGAAGGAGAAAATATCAGCACCTGTGGAGATGGGGGTGGAAATGGG							
AACAGCAAAACACGTTGCAATGAAGAGTGAAAGG-GGATCAGGAGGAATTAT-CAG-CACTGGTGGGGATGGG							
6220	6230	6240	6250	6260	6270	6280	
80	90	100	110	120	130	140	
GCACCCATGTCCTTCGATATTGATCTGTAGTAGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG							
GCACCCATGTCCTTCGATATTGATCTGTAGTAGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG							
6290	6300	6310	6320	6330	6340	6350	
150	160	170	180	190	200	210	
TACGTGTCTGGAAAGGAGCAACCAACACTCTATTTGTGCATCAGATGCTAAAGCATATGATAACAGAGGTAC							
TACGTGTCTGGAAAGGAGCAACCAACACTCTATTTGTGCATCAGATGCTAAAGCATATGATAACAGAGGTAC							
6360	6370	6380	6390	6400	6410	6420	6430
220	230	240	250	260	270	280	
ATAATGTTTGCGGCCACACATGCCCTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA							
ATAATGTTTGCGGCCACACAGCTGTGTACCCACAGACCCCAACCAAGAAGTAGAATTGGTAAATGTGA							
6440	6450	6460	6470	6480	6490	6500	
290	300	310	320	330	340	350	360
CAGAAATTTAACATGTCGAAAGATGACATGGTAGAACAGATGCGATGAGGATATAATCAGTTATGGGATC							
CAGAAATTTAACATGTCGAAAGATGACATGGTAGAACAGATGCGATGAGGATATAATCAGTTATGGGATC							
6510	6520	6530	6540	6550	6560	6570	
370	380	390	400	410	420	430	
AAAGGCTAAAGCCATTTAAATTAACCCCCACTCTGTGTTAGTTAAAGTGCAGTTGGGGATGCTA							
AAAGGCTAAAGCCATTTAAATTAACCCCCACTCTGTGTTACTTTAAATTGCACTGATTGGAGAAACTA							
6580	6590	6600	6610	6620	6630	6640	
440	450	460	470	480	490		
CTATTCACCAATGCTA-----TAATACCAATAGTAGTAGCGGGGAAATGATGATGGAGAAAGGAGAGATAA							
CTATTCACCAATGCTA-----TAATACCAATGCTAATAGCGAGGGAAACAATAAAGG-GA--GGAGAAATGA							
6650	6660	6670	6680	6690	6700	6710	
500	510	520	530	540	550	560	570
AAAATGCGTTTCAAATCGCAAAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTGTTTATAAAC							
AAAATGCGTTTCAAATCGCAAAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTGTTTATAAAC							
6720	6730	6740	6750	6760	6770	6780	

510 550 600 610 620 630 640
TTGATATACTTAACTGATAATGACTACCAAGCTATACTGGTACAAGTTGTAACACCTCAGTCATTACAC
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TTGATATACTTAACTGATAATGACTACCAAGCTATACTGGTACAAGTTGTAACACCTCAGTCATTACAC
6790 6800 6810 6820 6830 6840 6850 6860

650 660 670 680 690 700 710
AGGCCTGTCCAAAGGTTATCCCTTGAGCCAATTCCCATACTTATTGTGCCCGGCTGGTTTGCATTCTAA
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAGGTTTCTCCAAAGATATCCTTGTGAGCCAATTCCCATACTTATTGTGCCCGGCTGGTTTGCATTCTAA
6870 6880 6890 6900 6910 6920 6930

720 730 740 750 760 770 780
AATGTTAACTTAACTGAGGTTCTGATACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAA
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
AATGTTAACTTAACTGAGGTTCTGATACAGGACCATGTACAAATGTACACAGTACAATGTACACATGGAA
6940 6950 6960 6970 6980 6990 7000

790 800 810 820 830 840 850
TTAGGCCCCATGATCTGAACTCAACTCTGTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTG
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
TTAGGCCCCATGATCTGAACTCAACTCTGTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTG
7010 7020 7030 7040 7050 7060 7070

860 870 880 890 900 910 920 930
CCAAATTCTACAGACAAATGCTAAAAACATAATACTGACAGCTGAACCAATCTGTAGAAATTAAATTGTACAAGAC
||||| ||| ||||| ||||| ||||| ||||| |||||
AGAATTTCACCTGATCTGCTAAACACATCATAGTACATCTGAATCTGTACAAATTAAATTGTACAAGAC
7080 7090 7100 7110 7120 7130 7140

940 950 960 970 980 990 1000
CCACAAACAATACAAGAAAAGTATCGTATCCUAGAGGGGACCCAGGAGAGCATTTGTTACAATAGGAA--
||||| ||| ||||| ||| ||| ||||| ||||| |||||
CCAACTTACAAATAAAAGAAAAGATACATAT--AG----GACCAGGGAGAGCATTTTATACAACACAAAAATA
7150 7160 7170 7180 7190 7200 7210

1010 1020 1030 1040 1050 1060 1070
-AATAGGAAATATGAAACACATTGTAAACATTAGTAGAGCAAATGCAATGCCACTTAAACAGATAG
||||| ||| ||||| ||| ||| ||||| |||||
TAATAGGAAATATGAAACACATTGTAAACATTAGTAGAGCAAATGGAATGACACTTTAAGACAGATAG
7220 7230 7240 7250 7260 7270 7280

1080 1090 1100 1110 1120 1130 1140
CTAGCAAAATTAAAGAACAAATTGAAATAATAAAACAATACTTTAAGCAATCCTCAGGAGGGGACCCAG
||||| ||| ||||| ||| ||| ||||| |||||
TTAGCAAAATTAAAGAACAAATTGAAATAAAACAATACTGTTAATCAATCCTCAGGAGGGGACCCAG
7290 7300 7310 7320 7330 7340 7350

1150 1160 1170 1180 1190 1200 1210
AAATGTTAGGACACATTGAAATGAGGGGAATTTTCTACTGTAATTCAACACAACGTGTTAATAGTA
||||| ||| ||||| ||| ||| ||||| |||||
AAATGTTAGGACACATTGAAATGAGGGGAATTTTCTACTGTAATACATCACCACGTGTTAATAGTA
7360 7370 7380 7390 7400 7410 7420

1220 1230 1240 1250 1260 1270 1280
CTTGG---TTTGTAGTACCTGG---AGTACTGAGGGTCAAATAACACTGAAAGGAAGTGACACAATCACAC
||||| ||| ||||| ||| ||| ||||| |||||
CTTGGAAATGAACTTAACTTGGAAATAACTACAGGGTCAAATAACAAAT-----ATCACAC
7430 7440 7450 7460 7470 7480

1290 1300 1310 1320 1330 1340 1350
GGGATGCGAAATAAAACAATTTATAACATGTGGAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCA
||||| ||| ||||| ||| ||| ||||| |||||
TTGAGTGGAAATGAAACAAATTAAACATGTGGAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATG
7490 7500 7510 7520 7530 7540 7550

1320 1370 1380 1390 1400 1410 1420
 GCGGACCAATTAGATTTTCACTAAATATTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACA---AC-
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AACATCAAAATTAGATTTTCACTAAATATTACAGGGCTACTATTAAACAAGAGATGGTGGTAAGGACACGGACA
 7550 7570 7580 7590 7600 7610 7620
 1430 1440 1450 1460 1470 1480 1490
 ---ATGGGTCCTCAGAATTTCAGAACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAAAT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CGAACACACACCOLAGATCTTCAGAACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAAAT
 7630 7640 7650 7660 7670 7680 7690 7700
 1500 1510 1520 1530 1540 1550 1560
 ATAAAGTAGTAAACATTGAACCATTAGCAGTACACCCACCAAGGCAAAGAGAAAGAGTGGTGCAGAGAGAAA
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ATAAAGTAGTAAACATTGAACCATTAGCAGTACACCCACCAAGGCAAAGAGAAAGAGTGGTGCAGAGAGAAA
 7710 7720 7730 7740 7750 7760 7770
 1570 1580 1590 1600 1610 1620 1630
 AAAAGGGCACTTGATAGGAGCTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGCGCAGCGT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AAAGGGCAAGG---ATGGGAGCTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGCGCAGCGT
 7780 7790 7800 7810 7820 7830 7840
 1640 1650 1660 1670 1680 1690 1700
 CAATHACCUAGACGTTACAGGCCAACAAATTATGTCAGGTTAGCAGCAGCAGAACAAATTGCTGAGGG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CAGTGGGGCTGAGGTTACAGGCCAACAAATTATGTCAGGTTAGCAGCAGCAGAACAAATTGCTGAGGG
 7850 7860 7870 7880 7890 7900 7910
 1710 1720 1730 1740 1750 1760 1770
 CTATTGACCGCAGAGCATCTTCTGCACTCACAGTCTGGGGCATCAAGCAGCTCAGGCAAGAACATCCTGG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CCATTAAGGGCGAACAGCATATGTGCAACTCACAGTCTGGGGCATCAAGCAGCTCAGGCAAGAGTCTGG
 7920 7930 7940 7950 7960 7970 7980
 1780 1790 1800 1810 1820 1830 1840 1850
 CTGTAAAGATACCTAAAGGATCAACAGCTCTGGGATTTGGGTTGCTCTGGAAAACCTCATTTGCACCA
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTGTAAAGATACCTAAAGGATCAACAGCTCTGGGTTTGGGTTGCTCTGGAAAACCTCATTTGCACCA
 7990 8000 8010 8020 8030 8040 8050
 1960 1970 1980 1990 2000 2010 2020
 CTCGTTCTGAGTGGAAATGCTAGTGGAACTAAATCTCTGGAACAGATTTGGAAATAACATGACCTGGATGG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTACCTGCTCTGGAAATGCTAGTGGAACTAAATCTCTGGATGATATTGGAAATAACATGACCTGGATGC
 8060 8070 8080 8090 8100 8110 8120
 1930 1940 1950 1960 1970 1980 1990
 AGTGGGACGAGAAATTAAACATTACACAAAGCTTAATACATTCCCTTAATTGAAGAATCGCAAAACCGAGCAAG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ASTGGAAAGAGAAATTGGAACTTACACAAAGCTTAATATACTCATTACTAGAAAAATCGCAAAACCCAAACAAAG
 8130 8140 8150 8160 8170 8180 8190 8200
 2000 2010 2020 2030 2040 2050 2060
 AAAAGGATCAACAAATTGGAAATTAGATTAATGGGCAAGTTGTGGATTGGTTAACATAACAAATT
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 AAAAGGATCAACAAATTGGGATTAATGGGCAAGTTGTGGATTGGTTAACATAACAAATT
 8210 8220 8230 8240 8250 8260 8270
 2070 2080 2090 2100 2110 2120 2130
 GGCTTCTGGATATAGAAATGATATGATACTAGGAGGCTTGCTAGGTTAAGAATAGTTTGCTGTAC
 ||| ||| ||| ||| ||| ||| ||| |||
 GGCTTCTGGATATAGAAATGATATGATACTAGGAGGCTTGCTAGGTTAAGAATAGTTTGCTGTAC
 8280 8290 8300 8310 8320 8330 8340

2140 2150 2160 2170 2180 2190 2200 2210
 TTTCATTAAGATACTTAAAGGATATTCAACCATTATCGTTAGACCCACCTCCCAACCCCCGAGGG
 TTTCTATACTTATGTTTGCAGACCCGCCCCCAGTTCCGAGGG
 8350 8360 8370 8380 8390 8400 8410
 2220 2230 2240 2250 2260 2270 2280
 GAGCGGAGAAGGAAAGGAATAAGAAGAAGGTGGAGAGAGACAGAGACAGATCCATTGATTAGTGA
 GACCCGAGAAGGAGGAAAGAACAGAAGAAGGTGGAGAGAGACAGAGACACATCCGGTCGATTAGTGC
 8420 8430 8440 8450 8460 8470 8480
 2290 2300 2310 2320 2330 2340 2350
 ACAGGATCCCTAAUACTTATCTGGGAGCGATCTGGGGAGCCTTGTGCCTCTTCAGCTACCACCGCTTGAGAGAC
 ATGCAATTCTTAAACATTATCTGGGAGCGACCTGGGGAGCC-TGTTCCTCTTCAGCTACCACACACAGAC
 8490 8500 8510 8520 8530 8540 8550
 2360 2370 2380 2390 2400 2410 2420
 TTACTCTTAACTTAACGAGGATTCTGGAACTTCTGGGACGCGCAGGGGGTGGGAAGGCCCTCAAATATTGGTGG
 TTACTCTTAACTTAACGAGGATTCTGGAACTTCTGGGACGCGCAGGGGGTGGGAAGTCCTCAAATATTGGTGG
 8560 8570 8580 8590 8600 8610 8620
 2430 2440 2450 2460
 AATCTCTTAACTTATCTGGGAGCTCAAGAACTAAAG
 AATCTCTTAACTTATCTGGGAGCTCAAGAACTAAAG
 8630 8640 8650 8660 X

3. KUNZ-158-UL32 5'ED

HIVCDC42 Human immunodeficiency virus type 1, isolate CDC-4

LOCUS	HIVCDC42	3373 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate CDC-451, tat, rev, env and nef genes.			
ACCESSION	M17137			
KEYWORDS	env gene; tat gene.			
SEGMENT	2 of 2			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate CDC-451, unintegrated circular proviral DNA.			
ORGANISM	Human immunodeficiency virus type 1			
	Viridae; ss-RNA enveloped viruses; Retroviridae;			
	Lentivirinae.			
REFERENCE	i (bases 1 to 3373)			
AUTHORS	Dessai, S. M., Kalaynaraman, V. S., Casey, J. M., Srinivasan, A., Andersen, P. R. and Devare, S. G.			
TITLE	Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences			
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 83, 8380-8384 (1986)			
STANDARD	full staff_review			
COMMENT	Kindly submitted in computer-readable form by [1]. The normal start codon for the nef gene is not present; the ATG at 3142 may serve this role.			
FEATURES	from	to/span	description	
pept	1	52	vpr protein, partial (AA at 2)	
pept	88	312	tat protein, exon 2 (first expressed exon)	
	2687	2757	tat protein, exon 3 (AA at 2668)	
pept	227	302	rev protein, exon 2 (first expressed exon)	
	2687	2341	rev protein, exon 3 (AA at 2669)	
pept, ps	315	555	vpu protein (in-frame stops at bases 451 and 484)	
pept	477	3053	envelope polyprotein	
pept, ps	1	3373	genomic mRNA	

pre-mag < 1 > 3373 tat, rev, nef subgenomic mRNA
 IVS < 1 > 34 tat, rev, nef subgenomic mRNA intron 1
 IVS 303 2585 tat cds intron 2
 IVS 303 2586 rev cds intron 2
 IVS 303 2586 tat, rev, nef subgenomic mRNA intron 2
 BASE COUNT 1174 a 526 c 809 g 805 t
 ORIGIN 1 bp upstream of EcoRI site; about 3.6 kb after segment 1.
 Initial Score = 932 Optimized Score = 2209 Significance = 0.00
 Residue Identity = 89% Matches = 2241 Mismatches = 196
 Gaps = 50 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
AAGACGAAAGAACATTTGCAATGGAACTGAAAGGAAATAATCAGCACTTGTGGAGATGGGGGTGGAAAATGGG							
AAGAAGAAAGAACAAATGGCAATGGAAAGGAAAGG-GGATCAGGAAGAATTGT-CAG-CACTTGTGGAGATGGG							
X 470 480 490 500 510 520 530							
80 90 100 110 120 130 140							
GCACCATGCTTCCTTGTGATCTGTAAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG							
GCACCATGCTTCCTTGTGATCTGTAAGTGCTGCAGCAAACCTTGTGGGTACAGTCTATTATGGGG							
540 550 560 570 580 590 600							
150 160 170 180 190 200 210							
TACCGTATAAGGAAAGGAAAGAACACACTCTA1TTTGTGATCAGATGCTAAAGCATATGATAACAGAGGTAC							
TACCGTATAAGGAAAGGAAAGAACACACTCTA1TTTGTGATCAGATGCTAAAGCATATGATAACAGAGGTAC							
610 620 630 640 650 660 670							
220 230 240 250 260 270 280							
ATAATGTTGGGGCACACATGCCGTGTACCCACAGACCCCCACAAGAAGTAGTATTGGTAAATGTGA							
ATAATGTTGGGGCACACATGCCGTGTACCCACAAACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA							
680 690 700 710 720 730 740							
230 300 310 320 330 340 350 360							
CAGAYAAATTCACATTTGGAAAGAAATGACATGTTAGAACAGATGCATGAGGGATAATCAGTTATGGGATC							
CAGAAAGAAATTTAACATTTGGAAAGAAATGACATGGTAGAACAGATGCATGAGGGATAATCAGCTTATGGGATC							
750 760 770 780 790 800 810							
370 380 390 400 410 420 430							
AAAGGCTTAAGGCACTGTTAAATTAACCCCACTCTGTGTTAGTTAAAGTGCACTGATTTGGGAATGCTA							
AAAGGCTTAAGGCACTGTTAAATTAACCCCACTCTGTGTTACTTTAAATTGCACTGATTTGAATACTAATA							
820 830 840 850 860 870 880 890							
440 450 460 470 480 490 500							
CTAAATACCAATACTA--GTAATACCAATAGTATGATGCGGGGAAATGATGATGGAGAAAGGGAGAGATAAAAAA							
ATACATCTAAATACTACTGAACTATUAATAATAAGTAGTTGGGAACAAACG--GGGTAAAGGGAGAAATGAGAAA							
900 910 920 930 940 950 960							
510 520 530 540 550 560 570							
CTGTCTTTCAATATGCGACACAGATAAGGGTAAGGTGCAGAAAGAATATGCATTTTTATAAAACTTGA							
CTGTCTTTCAATATGACACAGATAAGGGTAAGGTGCAGAGAGAAATATGCATTGTTTATAAAACTTGA							
970 980 990 1000 1010 1020 1030							
580 590 600 610 620 630							
TATAATACCAATACTA-----ATAATGATACCACTACCAGC-----TATACGTTGACAAGTTGTAACACCTC							
TGTAGAAACCAATAGATGATAATAATAACTACCAACACACCAAAATAGGTTGATAAAATTGTAACACCTC							
1040 1050 1060 1070 1080 1090 1100							

640 650 660 670 680 690 700
AGTCATTACACAGGCGTGTCAAAAGGTA TCCCTTGAGCCAATTCCCATAACATTATTGTGCCCGGGCTGGTT
1110 1120 1130 1140 1150 1160 1170

710 720 730 740 750 760 770
TGCAATTCTAAAGATGTTAATAGACGTTCAATGAAACAGGACCATGTACAAATGTCAGCACAGTACAATG
1180 1190 1200 1210 1220 1230 1240

780 790 800 810 820 830 840
TACACATGGAATTAGGUCCAGTATCAACTCAACTGCTTTGAATGGCAGTCTAGCAGAAGAAGAGGTTAGT
1250 1260 1270 1280 1290 1300 1310 1320

850 860 870 880 890 900 910
AATTAGATCTGAAATTTCAACAGACAATGCTAAAACCATAATAGTACAGCTGAACCAATCTGTAGAAATTAA
1330 1340 1350 1360 1370 1380 1390

920 930 940 950 960 970 980 990
TTGTACAAAGACCCAAACAAACATAACAGAAAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTGTTAC
1400 1410 1420 1430 1440 1450

1000 1010 1020 1030 1040 1050 1060
AATAAGGAAAT--TAAAGAATATAGACAAAGAACATTTGTAACATTAGTAGAGCAAAATGCAATGCCACTT
1460 1470 1480 1490 1500 1510 1520 1530

1070 1080 1090 1100 1110 1120 1130
AAAACAGAACTGAAATTAAAGAGAACATTTGGAAATAATAAAACAATAATCTTAAAGCAATCCTCAGG
1540 1550 1560 1570 1580 1590 1600

1140 1150 1160 1170 1180 1190 1200
AGGAAACCCAAATTGTAATGACAGTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCAACACAACT
1610 1620 1630 1640 1650 1660 1670

1210 1220 1230 1240 1250 1260
GTTTAATAAGTACTG---GTT---TAATACTACTGGAG---TACT-GAAGGGTCAAATA-ACACTGAA
1680 1690 1700 1710 1720 1730 1740

1270 1280 1290 1300 1310 1320 1330
GGAACATGACAAATCAUACTCCCATGCAGAACATAAAACAATTAAACATGTGGCAGGAAGTAGGAAAAGCA
1750 1760 1770 1780 1790 1800

1340 1350 1360 1370 1380 1390 1400
ATGTTTGCGCTCGCA TCAAGGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAAGAGAT
1810 1820 1830 1840 1850 1860 1870 1880

1410 1420 1430 1440 1450 1460 1470
GGTGTAAACAAACAAATGGGTCAGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGT
1480 1490 1500 1510 1520 1530 1540 1550
GAATTTATATAAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAGGAAAGAGTG
1560 1570 1580 1590 1600 1610
GTCAGAGAGAAGAAAAAAGCAGTGGAA---TAGGAGCTTGTCTGGTTCTGGGAGCAGCAGGAAGC
1620 1630 1640 1650 1660 1670 1680 1690
ACTATGGCGCGACGGTCAATGACCGTACAGGCCAGACAATTATTGTCTGGTATAGTCAGCAGCAG
1700 1710 1720 1730 1740 1750 1760
AACATTTGCTAGGGTATTGAGGCGAACAGCATCTGTTGCAACTCACAGTCTGGGCATCAAGCAGCTC
1770 1780 1790 1800 1810 1820 1830
CAGGCAAGAACCTGGCTGGAGATAACCTAAAGGATCAACAGCTCCTGGGATTTGGGTTGCTCTGGA
1840 1850 1860 1870 1880 1890 1900
AAACTCATTTGCAACCACTGCTGGCTTGGAAATGCTAGTTGGAGTAATAAACTCTGGAACAGATTGGAAT
1910 1920 1930 1940 1950 1960 1970
AACATGACCTGGATGGAGTGGGACAGAGAAAATTAAACAATTACACAAGCTTAATACATTCTTAATTGAAGAA
1980 1990 2000 2010 2020 2030 2040 2050
TCGCAAAACCAAAAGAAAATGAAACAGGAACTATTGCAATTAGATAAGTGGCAAGTTGTGGACTTGG
2060 2070 2080 2090 2100 2110 2120
TTTAACATAACAAATTGGCTGGTATATAAAATATTCTATAATGATAGTAGGAGGGCTGGTAGGTTAAGA
2130 2140 2150 2160 2170 2180 2190
ATAGTTTTGGCTGGTATATAAAATATTCTATAATGATAGTAGGAGGGCTGGTAGGTTAAGA
2530 2540 2550 2560 2570 2580 2590 2600

2200 2210 2220 2230
 CTCGCAACCCCGT AGGGAGCCGAGTAGGCCGAAGGAATAGAAGA
 ::::::::::::::::::::: :::::::::::::::::::::
 CTCGCGAAACCCGAGGCGAACCGCAGGCCGAAGGAACCGAAGAAGGAGGTGGAGAGAGA
 2280 2290 2300 2310 2320 2330
 TCCATTGCGATTGCTGAGGATCTCTTAGGACTTATCTGGGACGATCTGCGGAGGCCCTGTGCCTCTTCAGCTA
 ::::::::::::::::::::: :::::::::::::::::::::
 TCCGACTCGATTATGCGATGGCTCTTAGGACTTGCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGCTA
 2760 2780 2770 2780 2790 2800 2810

 2340 2350 2360 2370 2380 2390 2400 2410
 CCACCGCTTGAGAGACTTACTCTTCATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGC
 ::::::::::::::::::::: :::::::::::::::::::::
 CCACCGCTTGAGAGACTTACTCTGATTTAGCGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGT
 2860 2880 2890 2900 2910 2920 2930 X

 2420 2430 2440 2450 2460
 CCTCAAATATTGCGAAATCTCTACAGTATTGGAGTCAGGAACCTAAAG
 ::::::::::::::::::::: :::::::::::::::::::::
 CCTCAAATATTGCGAAATCTCTACAGTATTGGAGTCAGGAACCTAAAG
 2890 2900 2910 2920 2930 X

4. KUNZ-158-CL32. SEQ

HIVPV22 Human immunodeficiency virus type 1, isolate PV22,

LOCUS HIVPV22 9770 bp ss-RNA **VRL** 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate PV22, complete genome (H9/HTLV-III proviral DNA).
ACCESSION K02033
KEYWORDS TAR protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; rev gene; reverse transcriptase; tat gene; trans-activator.
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate PV22 (from H9-derived family), proviral DNA.
ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
REFERENCE 1 (bases 1 to 3770; revised sequence, personal communication)
AUTHORS Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Leaky, L. A. and Capon, D. J.
TITLE Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus
JOURNAL Nature 313, 450-458 (1985)
STANDARD full staff_review
REFERENCE 2 (bases 2111 to 2112; revises [1])
AUTHORS Muesing, M. A.
JOURNAL Unpublished (1987) Whitehead Inst Cambridge, Mass
STANDARD full staff_review
COMMENT This sequence for a H9/HTLV-III virus was determined from one complete proviral clone [1]. Additionally, several cDNA clones of the viral RNA were sequenced for comparison with the entire proviral sequence. The differences between cDNA and proviral DNA are extensive and are listed in the Sites Table as variations. The authors believe that the variations may be due in part to different strains in the H9/HTLV-III cell line, because it was established by infection with material from several AIDS patients. With the addition of g at 2111, gag cds and pol cds are very close to those of HXB2, BRU, and related HIV viruses. For details and other references pertaining to Sites and Features, see the HIV reference entry.
FEATURES

from	to/span	description
2901	2937	gag polyprotein precursor

pept	2084	5141	pol polyprotein (NH2-terminus uncertain; AA at 2084)
pept	5086	5864	vif protein
pept	5604	5840	vpr protein
pept	5876	6090	tat protein, exon 2 (first expressed exon)
	8421	8486	tat protein, exon 3 (AA at 8422)
pept	8015	6090	rev protein, exon 2 (first expressed exon)
	8421	8895	rev protein, exon 3 (AA at 8423)
pept	6107	6352	vpu protein
pept	6287	6837	envelope polyprotein
pept	8839	9453	nef protein
pre-msg	464	9678	genomic mRNA
pre-msg	464	9678	tat, rev, nef subgenomic mRNA
IVS	753	5922	tat, rev, nef subgenomic mRNA intron 1
IVS	6091	8420	tat cds intron 2
IVS	6091	8420	rev cds intron 2
IVS	6091	8420	tat, rev, nef subgenomic mRNA intron 2
LTR	10	543	5' LTR
LTR	8126	5761	3' LTR
rpt	463	580	R repeat 5' copy
rpt	7581	9678	R repeat 3' copy
binding	366	385	Spi binding site III
binding	397	406	Spi binding site II
binding	408	417	Spi binding site I
binding	646	662	primer (Lys-tRNA) binding site
variant	510	510	a in provirus; g in cDNA [1]
variant	575	575	g in provirus; a in cDNA [1]
revision	2111	2112	gg in [2]; g in [1]
variant	5716	5716	g in provirus; a in cDNA [1]
variant	5892	5892	a in provirus; g in cDNA [1]
variant	6007	6007	c in provirus; t in cDNA [1]
variant	6047	6047	c in provirus; g in cDNA [1]
variant	6051	6051	c in provirus; a in cDNA [1]
variant	6055	6057	egg in provirus; gaa in cDNA [1]
variant	6108	6108	t in provirus; c in cDNA [1]
variant	6120	6120	a in provirus; c in cDNA [1]
variant	6125	6126	gc in provirus; gtaac in cDNA [1]
variant	6136	6136	a in provirus; c in cDNA [1]
variant	6235	6235	t in provirus; a in cDNA [1]
variant	6352	6352	g in provirus; a in cDNA [1]
variant	6760	6760	t in provirus; a in cDNA [1]
variant	7080	7080	c in provirus; t in cDNA [1]
variant	7100	7100	a in provirus; g in cDNA [1]
variant	7134	7135	ca in provirus; ac in cDNA [1]
variant	7183	7184	gt in provirus; aa in cDNA [1]
variant	7193	7193	a in provirus; g in cDNA [1]
variant	7284	7295	aa in provirus; gc in cDNA [1]
variant	7303	7303	a in provirus; c in cDNA [1]
variant	7511	7511	a in provirus [1]; c in cDNA [1]
variant	7533	7533	t in provirus [1]; a in cDNA [1]
variant	7586	7586	c in provirus [1]; t in cDNA [1]
variant	7648	7648	a in provirus [1]; g in cDNA [1]
variant	8139	8139	a in provirus; c in cDNA [1]
variant	8143	8143	t in provirus; c in cDNA [1]
variant	8222	8222	g in provirus; a in cDNA [1]
variant	8269	8269	a in provirus [1]; g in cDNA [1]
variant	8285	8285	g in provirus [1]; t in cDNA [1]
variant	8376	8376	a in provirus [1]; g in cDNA [1]
variant	8381	6381	a in provirus [1]; g in cDNA [1]
variant	8476	8476	a in provirus [1]; g in cDNA [1]
variant	8869	8858	a in provirus [1]; g in cDNA [1]
variant	8978	8978	c in provirus; t in cDNA [1]
variant	8990	8990	a in provirus; c in cDNA [1]
variant	8999	8999	c in provirus [1]; a in cDNA [1]
variant	9031	9031	a in provirus [1]; g in cDNA [1]
variant	9291	9291	t in provirus [1]; a in cDNA [1]

variant 9285 9295 g in provirus [1]; t in cDNA [1]
 variant 9303 9303 g in provirus [1]; a in cDNA [1]
 variant 9548 9548 g in provirus [1]; c in cDNA [1]
 signal 9654 9659 mRNA polyadenylation signal
 prov 10 9761 HIV-1 proviral DNA
 cell 1 9 human cellular DNA
 cell 9782 9770 human cellular DNA
 BASE COUNT 3436 a 1786 c 2376 g 2172 t
 ORIGIN 482 bp upstream of BglII site.

 Initial Score = 1877 Optimized Score = 2190 Significance = 0.00
 Residue Identity = 89% Matches = 2258 Mismatches = 153
 Gaps = 110 Conservative Substitutions = 0

 X 10 20 30 40 50 60
 AAGAG-CAG---AAGACAGTGGCAATGAGAGTGAACGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA
 || || || || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AATAGACAGGTAAATTGATAGACTAATAGAAAG-AGCAGAACAGTGGCAAT---GAGAGTGAAGGAGAA
 6220 6230 6240 6250 6260 6270 6280

 70 80 90 100 110 120
 AATGGGGGCAC-CATGCTCCTGGCATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ATATCAGCACTTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTGGGATGTTGATCTGTAGTGC
 6290 6300 6310 6320 6330 6340 6350

 130 140 150 160 170 180
 CACAG-TCTATTGTCGGGTAC---CT---GTGTTGGAA-----GGAAGCAA-CCACCA-CTCTATTTGTG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TACAGAAAAATTGTGCTTACAGCTATTGTCGGGTACCTGTGGAAGGAAGCAACCACACTCTATTTG
 6360 6370 6380 6390 6400 6410 6420

 190 200 210 220 230 240
 CATCAGATGCTAAAGCATATGATA---CAGAGG-TACATA----AT--GTTTGGGCCACACATGCCTG--T
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ---TACAGATGCTAAAGCATATGATAACAGAGGTACATAATGTTGGC-CACA--CATGCCT
 6430 6440 6450 6460 6470 6480

 250 260 270 280 290 300 310
 GTACCCCACAGA-CCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGGAAAAA-
 || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GTGTACCCCACAGAACCCAAACCA-CA-CAAGAAGTA---GTA-TTG-----GTAAAT-GTGACA---GAAAAT
 6490 6500 6510 6520 6530 6540

 320 330 340 350 360 370
 --ATACATGCTAAACAGATE-CATGAGGATAATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-T
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TTAAACAT-GTGGAA-AAATGACAT---GGTAGAA-CAGATGCATGAGGATATAA---TCAGTTATGGGAT
 6550 6560 6570 6580 6590 6600

 380 390 400 410 420 430 440
 AAAA---TTAACCCCCACTCTGTGTTAGTTAA-AGTGCAC TGATTTGG---GGAATGCTACTAAT---ACCAA
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CAAAGCCTAAACCCA---TGTGTAAAATTAACCCCCACTCTGTGTTAGTTAAAGTGC-ACTGATTTGAAGAA
 6610 6620 6630 6640 6650 6660 6670

 450 460 470 480 490 500 510
 TACTAGTAATACCAATAGTAGTACCGGGAAATGATGATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCAA
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TGATACTAATACCAATAGTAGTACCGGGAGAAATGATAATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCAA
 6680 6690 6700 6710 6720 6730 6740

 520 530 540 550 560 570 580
 TATCAGGACACGATATAAGAGGTAGGGTGCAGAAAGAAATATGCATTTTTATAAAACTTGATATAATACCAAT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TATCAGGACACGATATAAGAGGTAGGGTGCAGAAAGAAATATGCATTTTTATAAAACTTGATATAATACCAAT

6750	6760	6770	6780	6790	6800	6810	
590	600	610	620	630	640	650	
AGATAATGATACTACCGAGC1ATACTTGTGACAAGTTGTAACACACCTCAGTCATTACACAGGCCTGTCCAAAGGT							
6820	6830	6840	6850	6860	6870	6880	
660	670	680	690	700	710	720	730
ATCCCTTCAACCCGATTCGGCATACATTATTCGGCTGGTTTGCGATTCTAAAATGTAATAATAAGAC							
6890	6900	6910	6920	6930	6940	6950	6960
740	750	760	770	780	790	800	
GTTCAATEGAAAGGACCATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC							
6970	6980	6990	7000	7010	7020	7030	
810	820	830	840	850	860	870	
AACTCAACTCTGTTGAAATGCACTCTAGCAGAAGGAGAGGTTAGTAATTAGATCTGCCAATTCACAGACAA							
AACTCAACTCTGTTGAAATGCACTCTAGCAGAAGGAGAGGTTAGTAATTAGATCTGCCAATTCACAGACAA							
7040	7050	7060	7070	7080	7090	7100	
880	890	900	910	920	930	940	
TGCTAAACCAATAATGACGCTAACCAATCTCTAGAAATTAAATTGTACAAGACCCAACAACAATAACAG							
7110	7120	7130	7140	7150	7160	7170	
950	960	970	980	990	1000	1010	
AAAAAGTATCCATATCAGAGGGACCAAGGGAGAGCATTGTTACAATAGGAAAAATAGGAAATATGAGACA							
7180	7190	7200	7210	7220	7230	7240	
1020	1030	1040	1050	1060	1070	1080	1090
AGCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAAC							
7250	7260	7270	7280	7290	7300	7310	7320
1100	1110	1120	1130	1140	1150	1160	
ATTTGGAAATTAACAAATACTTAAAGCAATCCTCAGGGAGGGACCCAGAAATTGTAACGCACAGTT							
7330	7340	7350	7360	7370	7380	7390	
1170	1180	1190	1200	1210	1220	1230	
TAATTGTTGGAGGGAAATTCTACTGTAATTCAACACAACTGTTAATAGTACTTGGTTAATAGTACTTG							
7400	7410	7420	7430	7440	7450	7460	
1240	1250	1260	1270	1280	1290	1300	
GAGTACTGAGGGCTAACAACTGAGGGAAAGTGACACAACTCACACTCCCAGCAGAATAAAACAATTAT							
7470	7480	7490	7500	7510	7520	7530	
1310	1320	1330	1340	1350	1360	1370	
AAACCTTCTCTGGAAATAGGAAATGCAATGTTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA							
AAACCTTCTCTGGAAATAGGAAATGCAATGTTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA							

7540	7550	7560	7570	7580	7590	7600	
1380	1390	1400	1410	1420	1430	1440	1450
TATTACAGGCTCGTATAACAAAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGG							
TATTACAGGCTCGTATAACAAAGAGATGGTGGTAATAACAACAATGGTCCGAGATCTTCAGACCTGGAGG							
7610	7620	7630	7640	7650	7660	7670	7680
1460	1470	1480	1490	1500	1510	1520	
AGGAAGATAAGGGAGAATTGGAAAGTGAAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT							
AGGAAGATAAGGGAGAATTGGAAAGTGAAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT							
7690	7700	7710	7720	7730	7740	7750	
1530	1540	1550	1560	1570	1580	1590	
AGCAGCCAGCAAGGCAAAGAGAGAACAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCT							
AGCAGCCAGCAAGGCAAAGAGAGAACAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCT							
7760	7770	7780	7790	7800	7810	7820	
1600	1610	1620	1630	1640	1650	1660	
TGGGTTCTTGGGAGCACAGAAACCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT							
TGGGTTCTTGGGAGCACAGAAACCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT							
7830	7840	7850	7860	7870	7880	7890	
1670	1680	1690	1700	1710	1720	1730	
ATTCCTCTGTTAGTGAGCAGCAGAACATTGCTGAGGGCTATTGAGGGCGAACAGCATCTGTTGCAACT							
ATTCCTCTGTTAGTGAGCAGCAGAACATTGCTGAGGGCTATTGAGGGCGAACAGCATCTGTTGCAACT							
7930	7940	7950	7960				
1740	1750	1760	1770	1780	1790	1800	1810
CACAGCTGGGECATCAAGCAGCTCCAGGCAAGAACCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCT							
CACAGCTGGGECATCAAGCAGCTCCAGGCAAGAACCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCT							
7970	7980	8000	8010	8020	8030	8040	
1820	1830	1840	1850	1860	1870	1880	
CCTGGGGATTGGGGTTGCTCTGGAAAACCTCATTTGCACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAA							
CCTGGGGATTGGGGTTGCTCTGGAAAACCTCATTTGCACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAA							
8050	8060	8070	8080	8090	8100	8110	
1890	1900	1910	1920	1930	1940	1950	
TAAATCTCTGGYACAGATTGGAAATAACATGACCTGGATGGAGTGGACAGAGAAATTAAACAATTACACAAG							
TAAATCTCTGGYACAGATTGGAAATAACATGACCTGGATGGAGTGGACAGAGAAATTAAACAATTACACAAG							
8120	8130	8140	8150	8160	8170	8180	
1960	1970	1980	1990	2000	2010	2020	
CTTAATACATTCTTAATTGAAAGAACATCGCAAAACAGCAAGAAAAAGAATGAACAAGAAATTATTGGAATTAGA							
CTTAATACATTCTTAATTGAAAGAACATCGCAAAACAGCAAGAAAAAGAATGAACAAGAAATTATTGGAATTAGA							
8190	8200	8210	8220	8230	8240	8250	
2030	2040	2050	2060	2070	2080	2090	
TAAATGGGCGAACATTGGAAATTGGTAAACATAACAAATTGGCTGTGGTATATAAAATTCTATAATGAT							
TAAATGGGCGAACATTGGAAATTGGTAAACATAACAAATTGGCTGTGGTATATAAAATTCTATAATGAT							
8260	8270	8280	8290	8300	8310	8320	
2100	2110	2120	2130	2140	2150	2160	2170
AGTACGGAACTTGTAAGTTAACTGATAGTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA							
AGTACGGAACTTGTAAGTTAACTGATAGTTCTGTACTTTCTATAGTAAATAGAGTTAGGCAGGGATA							

8330	8340	8350	8360	8370	8380	8390	8400
2180	2190	2200	2210	2220	2230	2240	
TTCACCATTTATGGTTTCAAGACCCACCTCCCAACCCCCGAGGGGACCCGACAGGCCGAAGGAATAAGAAGAAGA							
TTCACCATTTATGGTTTCAAGACCCACCTCCCAACCCCCGAGGGGACCCGACAGGCCGAAGGAATAAGAAGAAGA							
8410	8420	8430	8440	8450	8460	8470	
2250	2260	2270	2280	2290	2300	2310	
AGGTGAGAGAGAGAGAGAGACATCCATTGGATTAGTGAAACGGATCCTAGCACTTATCTGGGACGATCT							
AGATGGAGAGAGAGAGAGAGACATCCATTGGATTAGTGAAACGGATCCTAGCACTTATCTGGGACGATCT							
8480	8490	8500	8510	8520	8530	8540	
2320	2330	2340	2350	2360	2370	2380	
GCGGAGGCCCTTGATGCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC							
GCGGAGGCCCTTGATGCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC							
8550	8560	8570	8580	8590	8600	8610	
2390	2400	2410	2420	2430	2440	2450	
TTCTGGGAGGCAAGGGGGTGGAAGACCTGAAATATGGGTGGAATCTCCTACAGTATTGGAGTCAGGAACCAA							
TTCTGGGAGGCAAGGGGGTGGAAGACCTGAAATATGGGTGGAATCTCCTACAAATATTGGAGTCAGGAGCTAA							
8620	8630	8640	8650	8660	8670	8680	
X							
AG							
AG							
8690							

5. KUNZ-158-CL32, SEE

HIVBH102 Human immunodeficiency virus type 1, isolate BH10, USA

LOCUS HIVBH102 SS32 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate BH10, genome.
 ACCESSION M10954 KC2010 K02008 K02009
 KEYWORDS TAT region; acquired immune deficiency syndrome; env gene;
 gag gene; long terminal repeat; pol gene; polyprotein; provirus;
 reverse transcriptase; trans-activator.
 SEGMENT 2 of 2
 SOURCE Human immunodeficiency virus type 1 (HIV-1) proviral DNA clone
 BH10.
 ORGANISM Human immunodeficiency virus type 1
 VIRION ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
 REFERENCE [1] (bases 1 to 3832)
 AUTHORS Ratner,L., Haseltine,W., Patarca,R., Livak,K. J., Starcich,B.,
 Josephs,S. E., Doran,E. R., Rafalski,J. A., Whitehorn,E. A.,
 Baumleister,K., Ivanoff,L., Petteway,S. R. Jr., Pearson,M. L.,
 Lautenberger,J. A., Papas,T. S., Ghayeb,J., Chang,N. T.,
 Gullion,R. C. and Wong-Staal,F.
 TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III
 JOURNAL Nature 313, 277-284 (1985)
 STANDARD full staff_review
 COMMENT The BH10 sequence differs from BH8 and BH5 by 0.9% in the coding
 regions and 1.8% in the noncoding regions, and the authors of [1]
 believe that these are stable variants.
 The HTLV-III genome encodes at least seven proteins: gag, pol, env,
 tef, trs, 27K antigen and the sor 23K product. The 3' ORF
 (positions 8153-8773) is truncated in BH10 (stop codon at positions
 8522-8524), but reads through in BH8 and other sequences to yield
 what is now called the 27K antigen.
 The mechanism for env gene translation has not been elucidated; a

gag-pol fusion protein is possible; splicing or frameshift have not been ruled out. The viral protease would be determined by the region in question.

The Tat protein (trans-activator protein, approximately 14 kb) is an effector of an autostimulatory pathway through interaction with a positive control element, the trans-activating responsive sequence, TAR. Tat seems to be a transcriptional control molecule in HTLV-I, but is both that and a post-transcriptional regulatory molecule in HTLV-III. Deletion mutants in the tat gene are incapable of prolific replication and exhibit no cytopathic effects in T4+ cell lines.

In addition to the ~9.4 kb genomic mRNA, subgenomic mRNAs of 7.4, 6.5, 5.0, 4.3, 2.0 and 1.8 kb have been detected.

FEATURES	from	to/span	description
pept	112	1850	gag polyprotein precursor
pept	< 1407	4454	pol polyprotein (NH2-terminus uncertain; AA at 1407)
pept	4395	4977	vif protein
pept	4817	5153	vpr protein
pept	5139	5403	tat protein, exon 2 (first expressed exon)
	7734	7779	tat protein, exon 3 (AA at 7735)
pept	5323	5403	rev protein, exon 2 (first expressed exon)
pept	7734	8003	rev protein, exon 3 (AA at 7736)
pept	5420	5865	vpu protein
pept	6590	8150	envelope polyprotein
pept	8152	8523	nef protein, exon 3 (first expressed exon; premature termination)
pre-msg	< 1	> 8932	genomic mRNA
pre-msg	< 1	> 8932	tat, rev, nef subgenomic mRNA
IVS	66	5135	tat, rev, nef subgenomic mRNA intron 1
IVS	5404	7733	tat cds intron 2
IVS	5404	7733	rev cds intron 2
IVS	5404	7733	tat, rev, nef subgenomic mRNA intron 2
LTR	8441	> 8932	3' LTR
Rpt	8894	> 8932	R repeat 3' copy
BASE COUNT	3227	a 1562	c 2147 g 1976 t
ORIGIN	About 120 bp downstream from HIVBH101.		

Initial Score = 1872 Optimized Score = 2186 Significance = 0.00
Residue Identity = 89% Matches = 2255 Mismatches = 157
Gaps = 110 Conservative Substitutions = 0

X	10	20	30	40	50	60
AAGAAK-CAG---AAGAACAGTGGCAATGAGAGTGAAAGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGGTGGGA						
AATAGACAGCTTAATTGATAGACTAATAGAAAG-AGCAGAAGACAGTGGCAAT---GAGAGTGAAGGAGAA						
5530	5540	5550	5560	5570	5580	5590
	70	80	90	100	110	120
AATGGGGCAC-CATGCTCCTTGATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT						
ATATCAGCACTTGTGAGATGGAGATGGGGCACCATGCTCCTGGGATGTTGATGATCTGTAGTGC						
5600	5610	5620	5630	5640	5650	5660
	130	140	150	160	170	180
CACAG-TCTATTAATGGGTAC---CT---GTGTGGAA-----GGAAGCAA-CCACCA-CTCTATTTGTG						
TACAGAAANATTGTCGTCACAGTCTATTATGGGTACCTGTGTGGAAAGGAAGCAACCACACTCTATTTG						
5670	5680	5690	5700	5710	5720	5730
	190	200	210	220	230	240
CATGAGATTTAAAGCATATAATA---CAGAGG-TACATA---AT--GTTGGGCCACACATGCCTG--T						
---TAC---ATUGAGTCTAAAGCATATGATACAGAGGTACATAATGTTGGGC-CACA---CATGCT						
	5750	5760	5770	5780	5790	

250 260 270 280 290 300 310
GTACCCACAGA-CCCCCAACCAACGAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGGAAA-
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
GTGTACCCAAATACCCCCAAC-GA-CAAGAAGTA---GTA-TTG----GTAAAT-GTGACA---GAAAAT
5800 5810 5820 5830 5840 5850

320 330 340 350 360 370
-ATGACATCTGAGAACAGATG-CATGAGGATATAATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-T
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
TTTAYCAT-GTGGAA-ATATGACAT---GGTAGAA-CAGATGCATGAGGATATAA---TCAGTTATGGGAT
5860 5870 5880 5890 5900 5910

380 390 400 410 420 430 440
AAAGA--TTAACCCACACTCTGTGTTAATGATGCACTGATTGG---GGAATGCTACTAAT---ACCAA
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
CAAAGGCTAAACCCA-TGTTGTAATAACCCCACTCTGTGTTAGTTAAAGTGC-ACTGATTGAAGAA
5920 5930 5940 5950 5960 5970 5980

450 460 470 480 490 500 510
TACAGATAATGCCATAGTATAGCGGGGAAATGATGATGGAGAAAGGAGAGATAAAAAACTGCTCTTCAA
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
TGATGCTAAATGCCATAGTATAGCGGGGAAATGATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTCAA
5990 6000 6010 6020 6030 6040 6050

520 530 540 550 560 570 580
TATCGCGAACGATAGGAGGTAAAGGTGCAAGAAATATGCATTTTTTATAAAACTTGATATAATACCAAT
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
TATCGCGAACGATAGGAGGTAAAGGTGCAAGAAATATGCATTTTTTATAAAACTTGATATAATACCAAT
6030 6040 6050 6060 6070 6080 6090

590 600 610 620 630 640 650
AGATTAATG
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
AGATTAATG
6140 6150 6160 6170 6180 6190 6200

660 670 680 690 700 710 720 730
ATCUTTTCGGTAAATCCATACATTATGTCGCCCGGGCTGGTTTGCATTCTAAATGTAATAATAAGAC
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
ATCUTTTCGGTAAATCCATACATTGTCGCCCGGGCTGGTTTGCATTCTAAATGTAATAATAAGAC
6210 6220 6230 6240 6250 6260 6270

740 750 760 770 780 790 800
GTTGAAATGACCGACCATGCAATGTCAGCACAGTACAATGTACACATGGATTAGGCCAGTAGTATC
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
GTTGAAATGACCGACCATGCAATGTCAGCACAGTACAATGTACACATGGATTAGGCCAGTAGTATC
6270 6280 6290 6300 6310 6320 6330 6340

810 820 830 840 850 860 870
AACTGACATGCTTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
AACTGACATGCTTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
6350 6360 6370 6380 6390 6400 6410

880 890 900 910 920 930 940
TGCTAAACGATATAGTACACGTAACCAATCTGAGAAATTAAATTGTACAAGACCCAACAACAATACAAG
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
TGCTAAACGATATAGTACACGTAACCAATCTGAGAAATTAAATTGTACAAGACCCAACAACAATACAAG
6420 6430 6440 6450 6460 6470 6480 6490

950 960 970 980 990 1000 1010
AATGAGTATGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
AATGAGTATGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
6570 6580 6590 6600 6610 6620 6630 6640
6550 6560

1020 1030 1040 1050 1060 1070 1080 1090
 AGGCACTTGTACATTAAGAAGAAAATGCAATCCACTTTAACAGATAGCTAGCAAATTAGAGAAACA
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 AGGCACTTGTACATTAAGAAGAAAATGCAATCCACTTTAACAGATAGCTAGCAAATTAGAGAAACA
 6570 6580 6590 6600 6610 6620 6630

 1100 1110 1120 1130 1140 1150 1160
 ATTTGGAAATACTAAAGAAATATCTTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAACGCACAGTTT
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 ATTTGGAAATACTAAAGAAATATCTTTAAGCACTCCTCAGGAGGGACCCAGAAATTGTAACGCACAGTTT
 6640 6650 6660 6670 6680 6690 6700

 1170 1180 1190 1200 1210 1220 1230
 TAATGGTGAACGGGAAATTCTACTGTAATTCAACACAACACTGTTAACAGTACTTGTTAACAGTACTTG
 ||||| ||||| ||||| ||||| ||||| |||||
 TAATGGTGAACGGGAAATTCTACTGTAATTCAACACAACACTGTTAACAGTACTTGTTAACAGTACTTG
 6710 6720 6730 6740 6750 6760 6770

 1240 1250 1260 1270 1280 1290 1300
 GAGACTCTAAACGATGAACTGAAAGGAAGTGACACAAATCACACTCCCATGCAGAATAAAACAATTAT
 ||||| ||||| ||||| ||||| ||||| |||||
 GAGACTCTAAACGATGAACTGAAAGGAAGTGACACAAATCACCCCTCCCATGCAGAATAAAACAATTAT
 6780 6790 6800 6810 6820 6830 6840 6850

 1310 1320 1330 1340 1350 1360 1370
 AAACATTCAGGAAAGTAAAGAAAGCAATGATGATGCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA
 ||||| ||||| ||||| ||||| |||||
 AAACATTCAGGAAAGTAAAGAAAGCAATGATGATGCCCTCCCATCAGTGGACAAATTAGATGTTCATCAA
 6860 6870 6880 6890 6900 6910 6920

 1380 1390 1400 1410 1420 1430 1440 1450
 TATTCACAGGCTCTATTAAACAAAGATGGTGGTAATAACAAACAATGGGTCCGAGATCTTCAGACCTGGAGG
 ||||| ||||| ||||| |||||
 TATTCACAGGCTCTATTAAACAAAGATGGTGGTAATAAGCAACAAATGAGTCCGAGATCTTCAGACCTGGAGG
 6930 6940 6950 6960 6970 6980 6990

 1460 1470 1480 1490 1500 1510 1520
 AGGAAATACTGGGAAATTGGAGAAAGTCATTATATAAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
 ||||| ||||| ||||| |||||
 AGGAAATACTGGGAAATTGGAGAAAGTCATTATATAAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
 7000 7010 7020 7030 7040 7050 7060

 1530 1540 1550 1560 1570 1580 1590
 AGCGACGCGAAAGGAAAGGAAAGTGTGAGACAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCT
 ||||| ||||| |||||
 AGCGACGCGAAAGGAAAGGAAAGTGTGAGACAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCT
 7070 7080 7090 7100 7110 7120 7130

 1610 1620 1630 1640 1650 1660
 TGGCTCTCTGGCGACGAAAGTACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
 ||||| ||||| |||||
 TGGCTCTCTGGCGACGAAAGTACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
 7140 7150 7160 7170 7180 7190 7200 7210

 1670 1680 1690 1700 1710 1720 1730
 ATTTCTCTTAATGTCAGGAAAGAAACAAATTGCTGAGGGCTATTGAGGCAGCAACAGCATCTGTTGCAACT
 ||||| ||||| |||||
 ATTTCTCTTAATGTCAGGAAAGAAACAAATTGCTGAGGGCTATTGAGGCAGCAACAGCATCTGTTGCAACT
 7240 7250 7260 7270 7280

 1740 1750 1760 1770 1780 1790 1800 1810
 CACAGTCTTGTGAAATGAGGCTTCAAGCAAAATCCTGGCTGAGGAAAGATAACCTAAAGGATCAACAGCT
 ||||| |||||
 CACAGTCTTGTGAAATGAGGCTTCAAGCAAAATCCTGGCTGAGGAAAGATAACCTAAAGGATCAACAGCT
 7290 7300 7310 7320 7330 7340 7350

6. KUNZ: 150-A150-555

HIV-1A: Human Immunodeficiency Virus type 1, isolate HXB3.

LOOKING FOR: 158 bp ss-RNA VRL 15-JUN-1989
DEFINITION: avian leukoencephalomyelitis virus type 1, isolate HYB3, env region

ACCESSION # : A001000
 KEYWORD :
 SOURCE : Human immunodeficiency virus type 1 (HIV-1), isolate HXB3, proviral DNA
 ORIGIN : Human immunodeficiency virus type 1
 Myxoviridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirusinae
 REFERENCES : See refs 1 to 3155.
 AUTHORS : Chou, R., Blangstedt, K., Gordon, M., Conroy, R., Schaber, M., Kramer, R.,
 Li, J. C., Wright-Palmer, F. and Reddy, E. P.
 TITLE : HIV-1 env gene products synthesized in E. coli are recognized by
 antibodies present in the sera of AIDS patients
 JOURNAL : J. VIROL. 41, 3703-3716 (1985)
 STATUS/TYPE : Full length_review
 FEATURES : From 3155
 description
 pprte : 6 51 vpr protein, partial (AA at 1)
 pprte : 51 251 tat protein, exon 2 (first expressed exon)
 pprte : 251 2633 tat protein, exon 3 (AA at 2633)
 pprte : 2633 2666 rev protein, exon 2 (first expressed exon)
 pprte : 2666 2693 rev protein, exon 3 (AA at 2634)
 pprte : 2693 3153 vpu protein
 pprte : 3153 3248 envelope polyprotein
 pprte : 3248 3356 nef protein, partial
 pprte : 3356 3455 genomic mRNA
 pprte : 3455 3528 tat, rev, nef subgenomic mRNA
 IVS : 3528 3533 tat, rev, nef subgenomic mRNA intron 1
 IVS : 3533 3631 tat cds intron 2
 IVS : 3631 3634 rev cds intron 2
 IVS : 3634 3751 tat, rev, nef subgenomic mRNA intron 2
 BASIC COUNT : 1054 a 827 c 766 g 759 t
 ORIGIN : Elongate approx. 88 bp upstream of tat initiation codon.
 Initial score = 1073 Optimized Score = 2186 Significance = 0.00
 Residue Identity = 85% Matches = 2243 Mismatches = 168
 Gaps = 130 Conservative Substitutions = 0

X	10	20	30	40	50	60
AATGAGG-GAA	-	AGA-AAGGTTTCTTGTGGAGAA	-	ATATCAGCACTTGTTGGAGA	-	TGGGGGGTGGAA
AT	AT	AT	AT	AT	AT	AT
ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA
430	440	450	460	470	480	490

X	50	60	70	80	90	100	110	120
ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA
AT	AT	AT	AT	AT	AT	AT	AT	AT
ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA	-	ATGAGAGGAA
500	510	520	530	540	550	560		

X	130	140	150	160	170	180		
CAUCU-CUAC	-	CAUCU-CUAC	-	CAUCU-CUAC	-	CAUCU-CUAC		
AT	AT	AT	AT	AT	AT	AT		
CAUCU-CUAC	-	CAUCU-CUAC	-	CAUCU-CUAC	-	CAUCU-CUAC		
570	580	590	600	610	620	630		

X	190	200	210	220	230	240	250
CATGAGATATG	-	CATGAGATATG	-	CATGAGATATG	-	CATGAGATATG	-
AT	AT	AT	AT	AT	AT	AT	AT
CATGAGATATG	-	CATGAGATATG	-	CATGAGATATG	-	CATGAGATATG	-
640	650	660	670	680	690		

X	260	270	280	290	300	310	
AGCC-AGCA	-	AGCC-AGCA	-	AGCC-AGCA	-	AGCC-AGCA	
AT	AT	AT	AT	AT	AT	AT	
AGCC-AGCA	-	AGCC-AGCA	-	AGCC-AGCA	-	AGCC-AGCA	
700	710	720	730	740	750		

320 330 340 350 360 370 380
 CATGTTTACGATGAACTTCAATATACTAGT-TTATG-GCATCAAAGCCTAAAGCCATGTG-TAAAAA-
 390 400 410 420 430 440
 GATGAACTGGCTTCGATTTTAAAGTGCAGCTATTGG---GGAATGCTACTAAT---ACCAATACTA
 450 460 470 480 490 500 510
 ATGCTTGCGATGTTTAACTGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 520 530 540 550 560 570 580 590
 GCGGAGGAACTGAGATGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 600 610 620 630 640 650 660
 ATGCTTGCGATGTTTAACTGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 670 680 690 700 710 720 730
 TTTCAGGAACTGAGATGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 740 750 760 770 780 790 800
 ATGCTTGCGATGTTTAACTGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 810 820 830 840 850 860 870
 GAGGAGGAACTGAGATGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 880 890 900 910 920 930 940 950
 AAGGAGGAACTGAGATGATGATGAGAGAGATAAAAAACTGCTCTTCATATCA
 960 970 980 990 1000 1010 1020
 GATGAACTGGCTTCGATTTTAAACAGATAGCTAGCAAATTAAAGAGAACAAATTG
 1030 1040 1050 1060 1070 1080 1090
 ATGCTTGCGATGTTTAAACAGATAGCTAGCAAATTAAAGAGAACAAATTG
 1100 1110 1120 1130 1140 1150 1160 1170
 1180 1190 1200 1210 1220 1230 1240
 1250 1260 1270 1280 1290 1300 1310 1320
 1330 1340 1350 1360 1370 1380 1390
 1400 1410 1420 1430 1440 1450 1460
 1470 1480 1490 1500 1510 1520 1530

1130 1140 1150 1160
GCAAGATTTTCTTACGAACTTGAACTTCGAAAGGGACCCAGAAATTGTAACGCACAGTTTAATT
1140 1150 1160 1170
GATGTTGTTTAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACAGTTTAATT
1540 1550 1560 1570 1580 1590 1600

1170 1180 1190 1200 1210 1220 1230
CTGCCAGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACAGTTGGAGTA
1180 1190 1200 1210 1220 1230 1240
GTCGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACAGTTGGAGTA
1610 1620 1630 1640 1650 1660 1670 1680

1240 1250 1260 1270 1280 1290 1300 1310
CTTCGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTUCCATGCAGAATAAAACAATTTATAAAC
1250 1260 1270 1280 1290 1300 1310 1320
CTTCGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1320 1330 1340 1350 1360 1370 1380
TGTGCGGGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1330 1340 1350 1360 1370 1380 1390 1400
TGTGCGGGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1780 1790 1800 1810 1820

1400 1410 1420 1430 1440 1450
CAGACCTGGGAGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1410 1420 1430 1440 1450 1460 1470 1480
CAGACCTGGGAGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1460 1470 1480 1490 1500 1510 1520
ATACGAGGAGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1490 1500 1510 1520 1530 1540 1550 1560
ATACGAGGAGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTATAAAC
1560 1570 1580 1590 1600 1610 1620 1630
1530 1540 1550 1560 1570 1580 1590 1600
GCAAGATTTCTTACGAACTTGAACTTCGAAAGGGACCCAGAAATTGTAACGCACAGTTGGCTTGGGT
1540 1550 1560 1570 1580 1590 1600 1610
GCAAGATTTCTTACGAACTTGAACTTCGAAAGGGACCCAGAAATTGTAACGCACAGTTGGCTTGGGT
1970 1980 1990 2000 2010 2020 2030 2040

1600 1610 1620 1630 1640 1650 1660 1670
TGTGCGGGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
1610 1620 1630 1640 1650 1660 1670 1680
TGTGCGGGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
2020 2030 2040 2050 2060 2070 2080 2090

1680 1690 1700 1710 1720 1730 1740
CTTCGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
1690 1700 1710 1720 1730 1740 1750 1760
CTTCGGAACTTGTGAACTTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
2130 2140 2150 2160 2170 2180

1750 1760 1770 1780 1790 1800 1810
TCTGGGGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
1760 1770 1780 1790 1800 1810 1820 1830
TCTGGGGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
2180 2190 2200 2210 2220 2230 2240 2250

1870 1880 1890 1900 1910 1920 1930 1940
GGATTTGGGGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
1880 1890 1900 1910 1920 1930 1940 1950
GGATTTGGGGGGATTCGAGGGACCCAGAAATTGTAACGCACACTCCATGCAGAATAAAACAATTTGT
2260 2270 2280 2290 2300 2310 2320

7. KUNZ-158-CL32. SEQ

HIVNL43 Human immunodeficiency virus type 1, NYS/BRU (LAV-)

LOCUS HIVNL43 8703 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, NY5/BRU (LAV-1) recombinant clone pNL4-3.
 ACCESSION M18921
 KEYWORDS "
 SOURCE Human immunodeficiency virus type 1 (HIV-1), NY5/BRU (LAV-1) recombinant clone pNL4-3.
 ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
 REFERENCE J. Virol. 53: 1 to 9709
 AUTHORS Black, J. D., Bannister-White, A. J., Willey, R. L. and McCoy, J.
 JOURNAL J. Virol. 53 (1980).
 STANDARD Full-length cDNA

REFERENCE	S. (Chem. 11 to 19708)		
AUTHORS	Ariach, M., Brandelman, H. E., Koenig, S., Folks, T., Willey, R., Roberts, P., and Martin, M. A.		
TITLE	Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone		
JOURNAL	J. Virol. 51, 2841-2851 (1986)		
STANDARD	full sequence review		
REFERENCE	B (simply revisions of [1])		
AUTHORS	Buckler, C. S.		
JOURNAL	Lysogeny (1988)		
STANDARD	full sequence review		
COMMENT	<p>Chimeric copy of sequence [1] kindly provided by Chuck Buckler, NIAID, Bethesda, MD. 04-JUNE-1988. The construction of pNL4-3 has been described in [1]. pNL4-3 is a recombinant (infectious) proviral clone that contains DNA from HIV isolates NY5 (5' half) and BRU (3' half). The site of recombination is the EcoRI site at positions 57043-57048.</p> <p>The length and sequence of the vpr coding region corresponds to that in the IIIB, SC, SF2, MAL and ELI isolates. The vpr coding region of these isolates is about 18 amino acid residues longer than the vpr coding region of the IIIB isolates. In HIVNL43, this length is due to a single base deletion (with respect to the IIIB's) at position 57040. The sequence at this position is "atttc" in HIVNL43 and "tttttc" in HIVXB2.</p> <p>The nef portion (BRU clone, sequenced by Wain-Hobson, et al. (Cell 40, 31-37 (1985))) and the BRU portion of the pNL4-3 recombinant clone were also used in clones from the same BRU isolate.</p> <p>There are two revisions reported in the FEATURES produced changes in codon usage frequencies. The revision at position 2421 changes one codon usage value from 'R' to 'G' in the pol coding region. The revisions at positions 2050-3000 changes three amino acid residues from 'WTRP' to 'WTRP' in the nef coding region.</p>		
FEATURES	From	to	span
pol	1119	3707	gap polyprotein
pept	1119	4085	pol polyprotein (NH2-terminus uncertain; AA at 20,?)
pept	1119	3711	vif protein
pept	3711	3834	vpr protein
pept	1119	4090	tat protein, exon 2 (first expressed exon)
pept	1119	4091	tat protein, exon 3 (AA at 8370)
pept	3711	3844	rev protein, exon 2 (first expressed exon)
pept	3711	3845	rev protein, exon 3 (AA at 8371)
pept	1119	3720	vpu protein
pept	1119	3721	envelope polyprotein
pept	1119	3722	nef protein
pept	1119	3723	genomic mRNA
pept	1119	3724	tat, rev, nef subgenomic mRNA
pept	1119	3725	tat, rev, nef mRNA intron 1
pept	1119	3726	tat cds intron 2
pept	1119	3727	rev cds intron 2
pept	1119	3728	tat, rev, nef mRNA intron 2
pept	1119	3729	5' LTR
pept	1119	3730	3' LTR
pept	1119	3731	R repeat 5' copy
pept	1119	3732	R repeat 3' copy
binding	1119	3733	Spi binding site III
binding	1119	3734	Spi binding site II
binding	1119	3735	Spi binding site I
binding	1119	3736	primed (Lys-tRNA) binding site
site	1119	3737	EcoRI site of recombination
start	1119	3738	HIV-1 isolate NY5 DNA end/HIV-1 isolate LAV DNA start
start	1119	3739	at in [31] tg in [1]
start	1119	3740	g in IS116 c in [1]
start	1119	3741	c in IS116 a in [1]

Initial window = 172, Utilized Score = 2181 Significance = 0.00
 Residual heterogeneity = 181.0 Mismatches = 187
 Gaps = 0 Conservative Substitutions = 0

Start	End	Step	Score	Start	End	Step	Score
20	30	40	50	60	70		
610	620	630	640	650	660	670	680
690	700	710	720	730	740	750	760
770	780	790	800	810	820	830	840
850	860	870	880	890	900	910	920
930	940	950	960	970	980	990	1000
1010	1020	1030	1040	1050	1060	1070	1080
1090	1100	1110	1120	1130	1140	1150	1160
1170	1180	1190	1200	1210	1220	1230	1240
1250	1260	1270	1280	1290	1300	1310	1320
1330	1340	1350	1360	1370	1380	1390	1400
1410	1420	1430	1440	1450	1460	1470	1480
1490	1500	1510	1520	1530	1540	1550	1560
1570	1580	1590	1600	1610	1620	1630	1640
1650	1660	1670	1680	1690	1700	1710	1720
1730	1740	1750	1760	1770	1780	1790	1800
1810	1820	1830	1840	1850	1860	1870	1880
1890	1900	1910	1920	1930	1940	1950	1960
1970	1980	1990	2000	2010	2020	2030	2040
2050	2060	2070	2080	2090	2100	2110	2120
2130	2140	2150	2160	2170	2180	2190	2200
2210	2220	2230	2240	2250	2260	2270	2280
2290	2300	2310	2320	2330	2340	2350	2360
2370	2380	2390	2400	2410	2420	2430	2440
2450	2460	2470	2480	2490	2500	2510	2520
2530	2540	2550	2560	2570	2580	2590	2600
2610	2620	2630	2640	2650	2660	2670	2680
2690	2700	2710	2720	2730	2740	2750	2760
2770	2780	2790	2800	2810	2820	2830	2840
2850	2860	2870	2880	2890	2900	2910	2920
2930	2940	2950	2960	2970	2980	2990	3000
3010	3020	3030	3040	3050	3060	3070	3080
3090	3100	3110	3120	3130	3140	3150	3160
3170	3180	3190	3200	3210	3220	3230	3240
3250	3260	3270	3280	3290	3300	3310	3320
3330	3340	3350	3360	3370	3380	3390	3400
3410	3420	3430	3440	3450	3460	3470	3480
3490	3500	3510	3520	3530	3540	3550	3560
3570	3580	3590	3600	3610	3620	3630	3640
3650	3660	3670	3680	3690	3700	3710	3720
3730	3740	3750	3760	3770	3780	3790	3800
3810	3820	3830	3840	3850	3860	3870	3880
3890	3900	3910	3920	3930	3940	3950	3960
3970	3980	3990	4000	4010	4020	4030	4040
4050	4060	4070	4080	4090	4100	4110	4120
4130	4140	4150	4160	4170	4180	4190	4200
4210	4220	4230	4240	4250	4260	4270	4280
4290	4300	4310	4320	4330	4340	4350	4360
4370	4380	4390	4400	4410	4420	4430	4440
4450	4460	4470	4480	4490	4500	4510	4520
4530	4540	4550	4560	4570	4580	4590	4600
4610	4620	4630	4640	4650	4660	4670	4680
4690	4700	4710	4720	4730	4740	4750	4760
4770	4780	4790	4800	4810	4820	4830	4840
4850	4860	4870	4880	4890	4900	4910	4920
4930	4940	4950	4960	4970	4980	4990	5000
5010	5020	5030	5040	5050	5060	5070	5080
5090	5100	5110	5120	5130	5140	5150	5160
5170	5180	5190	5200	5210	5220	5230	5240
5250	5260	5270	5280	5290	5300	5310	5320
5330	5340	5350	5360	5370	5380	5390	5400
5410	5420	5430	5440	5450	5460	5470	5480
5490	5500	5510	5520	5530	5540	5550	5560
5570	5580	5590	5600	5610	5620	5630	5640
5650	5660	5670	5680	5690	5700	5710	5720
5730	5740	5750	5760	5770	5780	5790	5800
5810	5820	5830	5840	5850	5860	5870	5880
5890	5900	5910	5920	5930	5940	5950	5960
5970	5980	5990	6000	6010	6020	6030	6040
6050	6060	6070	6080	6090	6100	6110	6120
6130	6140	6150	6160	6170	6180	6190	6200
6210	6220	6230	6240	6250	6260	6270	6280
6290	6300	6310	6320	6330	6340	6350	6360
6370	6380	6390	6400	6410	6420	6430	6440
6450	6460	6470	6480	6490	6500	6510	6520
6530	6540	6550	6560	6570	6580	6590	6600
6610	6620	6630	6640	6650	6660	6670	6680
6690	6700	6710	6720	6730	6740	6750	6760
6770	6780	6790	6800	6810	6820	6830	6840
6850	6860	6870	6880	6890	6900	6910	6920
6930	6940	6950	6960	6970	6980	6990	7000
7010	7020	7030	7040	7050	7060	7070	7080
7090	7100	7110	7120	7130	7140	7150	7160
7170	7180	7190	7200	7210	7220	7230	7240
7250	7260	7270	7280	7290	7300	7310	7320
7330	7340	7350	7360	7370	7380	7390	7400
7410	7420	7430	7440	7450	7460	7470	7480
7490	7500	7510	7520	7530	7540	7550	7560
7570	7580	7590	7600	7610	7620	7630	7640
7650	7660	7670	7680	7690	7700	7710	7720
7730	7740	7750	7760	7770	7780	7790	7800
7810	7820	7830	7840	7850	7860	7870	7880
7890	7900	7910	7920	7930	7940	7950	7960
7970	7980	7990	8000	8010	8020	8030	8040
8050	8060	8070	8080	8090	8100	8110	8120
8130	8140	8150	8160	8170	8180	8190	8200
8210	8220	8230	8240	8250	8260	8270	8280
8290	8300	8310	8320	8330	8340	8350	8360
8370	8380	8390	8400	8410	8420	8430	8440
8450	8460	8470	8480	8490	8500	8510	8520
8530	8540	8550	8560	8570	8580	8590	8600
8610	8620	8630	8640	8650	8660	8670	8680
8690	8700	8710	8720	8730	8740	8750	8760
8770	8780	8790	8800	8810	8820	8830	8840
8850	8860	8870	8880	8890	8900	8910	8920
8930	8940	8950	8960	8970	8980	8990	9000
9010	9020	9030	9040	9050	9060	9070	9080
9090	9100	9110	9120	9130	9140	9150	9160
9170	9180	9190	9200	9210	9220	9230	9240
9250	9260	9270	9280	9290	9300	9310	9320
9330	9340	9350	9360	9370	9380	9390	9400
9410	9420	9430	9440	9450	9460	9470	9480
9490	9500	9510	9520	9530	9540	9550	9560
9570	9580	9590	9600	9610	9620	9630	9640
9650	9660	9670	9680	9690	9700	9710	9720
9730	9740	9750	9760	9770	9780	9790	9800
9810	9820	9830	9840	9850	9860	9870	9880
9890	9900	9910	9920	9930	9940	9950	9960
9970	9980	9990	10000	10010	10020	10030	10040

6780 6790 6800 6810 6820 6830 6840 6850
6860 6870 6880 6890 6900 6910 6920
6930 6940 6950 6960 6970 6980 6990
7000 7010 7020 7030 7040 7050 7060
7070 7080 7090 7100 7110 7120 7130
7140 7150 7160 7170 7180 7190 7200 7210
7220 7230 7240 7250 7260 7270 7280
7290 7300 7310 7320 7330 7340 7350
7360 7370 7380 7390 7400 7410 7420
7430 7440 7450 7460 7470 7480 7490
7500 7510 7520 7530 7540 7550 7560 7570
7580 7590 7600 7610 7620 7630 7640 7650
7660 7670 7680 7690 7700 7710 7720 7730
7740 7750 7760 7770 7780 7790 7800 7810
7820 7830 7840 7850 7860 7870 7880 7890
7890 7900 7910 7920 7930 7940 7950 7960
7970 7980 7990 8000 8010 8020 8030 8040
8050 8060 8070 8080 8090 8100 8110 8120
8130 8140 8150 8160 8170 8180 8190 8200
8210 8220 8230 8240 8250 8260 8270 8280
8290 8300 8310 8320 8330 8340 8350 8360
8370 8380 8390 8400 8410 8420 8430 8440
8450 8460 8470 8480 8490 8500 8510 8520
8530 8540 8550 8560 8570 8580 8590 8600
8610 8620 8630 8640 8650 8660 8670 8680
8690 8700 8710 8720 8730 8740 8750 8760
8770 8780 8790 8800 8810 8820 8830 8840
8850 8860 8870 8880 8890 8900 8910 8920
8930 8940 8950 8960 8970 8980 8990 9000
8990 9000 9010 9020 9030 9040 9050 9060
9070 9080 9090 9100 9110 9120 9130 9140
9150 9160 9170 9180 9190 9200 9210 9220
9230 9240 9250 9260 9270 9280 9290 9300
9310 9320 9330 9340 9350 9360 9370 9380
9390 9400 9410 9420 9430 9440 9450 9460
9470 9480 9490 9500 9510 9520 9530 9540
9550 9560 9570 9580 9590 9600 9610 9620
9630 9640 9650 9660 9670 9680 9690 9700
9710 9720 9730 9740 9750 9760 9770 9780
9790 9800 9810 9820 9830 9840 9850 9860
9870 9880 9890 9900 9910 9920 9930 9940
9950 9960 9970 9980 9990 9990 9990 9990

7580 7590 7600 7610 7620 7630 7640
 1520 1530 1540 1550 1560 1570 1580 1590 1600
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGG
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGG
 7650 7660 7670 7680 7690 7700 7710
 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670
 CGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGTTCTGGGAG
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGTTCTGGGAG
 7720 7730 7740 7750 7760 7770 7780
 1610 1620 1630 1640 1650 1660 1670 1680 1690 1700 1710 1720 1730 1740 1750
 CGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGTTCTGGGAG
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGTTCTGGGAG
 7790 7800 7810 7820 7830 7840 7850
 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960
 ATCCAGCAATGAGGCAAGATGCTGGTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGATTTGG
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGATTTGG
 7860 7870 7880 7890 7900 7910 7920 7930
 1760 1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGATTTGG
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGATTTGG
 7940 7950 7960 7970 7980 7990 8000
 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030
 CGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 8010 8020 8030 8040 8050 8060 8070
 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100 2110
 TTAACTGAGAATGAGGAAACGAAAGAAATGAAACAAGAATTATTGGAATTAGATAAAATGGGCAAGT
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 8050 8060 8070 8080 8090 8100 8110 8120 8130 8140
 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160 2170 2180
 TTAACTGAGAATGAGGAAACGAAAGAAATGAAACAAGAATTATTGGAATTAGATAAAATGGGCAAGT
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 8150 8160 8170 8180 8190 8200 8210
 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 GCGATGCTGGTAACTTGTGAAATTAATGAGTAAAAAGAGCAGTGGGAATAGGAGCTTGTCCTGGGAAAGT
 8240 8250 8260 8270 8280 8290
 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 2310 2320 2330 2340 2350

TTTTAGACGCGACGTCGGATCCGAGGGACCCGACAGGCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGA
 8370 8380 8390 8400 8410 8420 8430
 8440 8450 8460 8470 8480 8490 8500
 8510 8520 8530 8540 8550 8560 8570
 8580 8590 8600 8610 8620 8630 X

8. KJUZ-158-GLB, 839

HIV-1-HXB2 Human immunodeficiency virus type 1 (HXB2), complete genome

LOCUS HIV-1/HXB2 8719 bp ss-RNA
DEFINITION Human immunodeficiency virus type 1 (HXB2), complete genome; HIV-1/HTLV-III/LAV reference genome.
ACCESSION 104205
KEYWORDS TAR protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activator.
SOURCE HTLV-III/LAV (isolate HXB2) proviral DNA.
ORGANISM Human immunodeficiency virus type 1
VIRION ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
REFERENCE
AUTHORS Starcich,B., Rethner,L., Josephs,S. F., Okamoto,T., Gallo,R. C. and Wong-Staal,F.
TITLE Characterization of long terminal repeat sequences of HTLV-III
JOURNAL Science 237: 539-540 (1985)
STANDARD full staff_review
REFERENCE
AUTHORS R (bases 453 to 574; 577 to 9718)
TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III
JOURNAL Nature 313: 277-284 (1985)
STANDARD full staff_review
REFERENCE
AUTHORS Fisher,A. G., Rethner,L., Mitsuya,H., Marseille,L. M., Harper,M. E., Broder,S., Gallo,R. C. and Wong-Staal,F.
TITLE Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects
JOURNAL Science 233: 655-659 (1986)
STANDARD full staff_review
REFERENCE
AUTHORS Rethner,L., Fisher,A., Jegodzinski,L. L., Mitsuya,H., Liou,R.-S., Gallo,R. C. and Wong-Staal,F.
TITLE Complete nucleotide sequences of functional clones of the AIDS virus
JOURNAL AIDS Res. Hum. Retroviruses 3: 57-69 (1987)
STANDARD full staff_review
REFERENCE R (infect. tat mRNA and other transcript boundaries)

AUTHORS Arya,S. K., Guo,C., Josephs,S. F. and Wong-Staal,F.
TITLE Trans-activator gene of human T-lymphotropic virus type III (HTLV-III)
JOURNAL Science 233, 68-73 (1985)
STANDARD full staff_review
REFERENCE 6 (sites; tat mRNA)

AUTHORS Sodroski,J., Paterca,R., Rosen,C., Wong-Staal,F. and Haseltine,W.
TITLE Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III
JOURNAL Science 228, 74-77 (1985)
STANDARD full staff_review
REFERENCE 7 (sites; mRNA splice sites)

AUTHORS Rabson,A. E., Daugherty,D. F., Venkatesan,S., Boulukos,K. E., Benii,S. I., Folks,T. M., Feorino,P. and Martin,M.
TITLE Transcription of novel open reading frames of AIDS retrovirus during infection of lymphocytes
JOURNAL Science 229, 1388-1390 (1985)
STANDARD full staff_review
REFERENCE 8 (sites; 27K antigen cds)

AUTHORS Allen,J. S., Coligan,J. E., Lee,T.-H., McLane,M. F., Kanki,P. J., Broderman,J. E. and Essex,M.
TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients
JOURNAL Science 230, 810-813 (1985)
STANDARD full staff_review
REFERENCE 9 (sites; gp160 and gp120 coding sequences)

AUTHORS Allen,J. S., Coligan,J. E., Barin,F., McLane,M. F., Sodroski,J. G., Rosen,C. A., Haseltine,W. A., Lee,T. H. and Essex,M.
TITLE Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III
JOURNAL Science 228, 1091-1094 (1985)
STANDARD full staff_review
REFERENCE 10 (sites; regulatory sequences in the LTR)
AUTHORS Rosen,C. A., Sodroski,J. G. and Haseltine,W. A.
TITLE The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat
JOURNAL Cell 41, 813-823 (1985)
STANDARD full staff_review
REFERENCE 11 (review; bases 1 to 9718)
AUTHORS Van Beveren,C., Coffin,J. and Hughes,S.
TITLE Appendix B: HTLV-3 genome
JOURNAL (in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.); RNA Tumor Viruses, Second Edition, 2: 1102-1123 and 1147-1148; Cold Spring Harbor Laboratory, New York (1985)
STANDARD full staff_review
REFERENCE 12 (sites; trans-activator function and TAR sequence)
AUTHORS Rosen,C. A., Sodroski,J. G., Goh,W. C., Dayton,A. I., Lippke,J. and Haseltine,W. A.
TITLE Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III
JOURNAL Nature 319, 555-559 (1986)
STANDARD full staff_review
REFERENCE 13 (sites; pol coding sequence)
AUTHORS di Marzo Veronese,F., Cupeland,T. D., DeVico,A. L., Rahman,R., Grosszler,S., Gallo,R. C. and Sarngadharan,M. G.
TITLE Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV
JOURNAL Science 231, 1289-1291 (1986)
STANDARD full staff_review
REFERENCE 14 (sites; 23K sor gene product)
AUTHORS Wan,N. C., Frenchini,G., Wong-Staal,F., DuBois,G. C., Robey,W. G., Leutenberger,J. A. and Pepas,T. S.
TITLE Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera
JOURNAL Science 231, 1553-1555 (1986)
STANDARD full staff_review

REFERENCE	15 (sites; pol NH2-terminal region)
AUTHORS	Krammer,R. A., Scheber,M. D., Skalka,A. M., Ganguly,K., Wong-Staal,F. and Reddy,E. P.
TITLE	HTLV-III gag protein is processed in yeast cells by the virus pol protease
JOURNAL	Science 231, 1580-1584 (1986)
STANDARD	full staff_review
REFERENCE	16 (sites; sor 23K protein)
AUTHORS	Lee,T.-H., Galigan,J. E., Allan,J. S., McLane,M. F., Groopman,J. E. and Essex,M.
TITLE	A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses
JOURNAL	Science 231, 1546-1549 (1986)
STANDARD	full staff_review
REFERENCE	17 (sites; sor 23K protein)
AUTHORS	Sodroski,J., Goh,W. C., Rosen,C., Tartar,A., Portetelle,D., Rummel,A. and Haseltine,W.
TITLE	Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions
JOURNAL	Science 231, 1549-1553 (1986)
STANDARD	full staff_review
REFERENCE	18 (sites; Sp1 binding sites in the promoter region)
AUTHORS	Jones,K. A., Kadonaga,J. T., Luciw,P. A. and Tjian,R.
TITLE	Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1
JOURNAL	Science 232, 755-759 (1986)
STANDARD	full staff_review
REFERENCE	19 (sites; acceptor and donor splice sites for tat and 27K)
AUTHORS	Arya,S. K. and Gallo,R. C.
TITLE	Three novel genes of human T-lymphotropic virus type III: Immune reactivity of their products with sera from acquired immune deficiency syndrome patients
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 83, 2209-2213 (1986)
STANDARD	full staff_review
REFERENCE	20 (sites; deletion mutants in the tat gene)
AUTHORS	Dayton,A. I., Sodroski,J. G., Rosen,C. A., Goh,W. C. and Haseltine,W. A.
TITLE	The trans-activator gene of the human T cell lymphotropic virus type III is required for replication
JOURNAL	Cell 44, 541-547 (1986)
STANDARD	full staff_review
REFERENCE	21 (sites; env gene conserved/variable regions; separate entries)
AUTHORS	Williley,R., Rutledge,R. A., Dias,S., Folks,T., Theodore,T., Buckler,C. E. and Martin,M. A.
TITLE	Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome virus
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 83, 5038-5042 (1986)
STANDARD	full staff_review
REFERENCE	22 (sites; trs cds boundaries)
AUTHORS	Sodroski,J., Goh,W. C., Rosen,C., Dayton,A. I., Terwilliger,E. and Haseltine,W.
TITLE	A second post-transcriptional trans-activator gene required for HTLV-III replication
JOURNAL	Nature 321, 412-417 (1986)
STANDARD	full staff_review
REFERENCE	23 (sites; trs cds boundaries)
AUTHORS	Furberg,M. B., Jarret,R. F., Aldovini,A., Gallo,R. C. and Wong-Staal,F.
TITLE	HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA
JOURNAL	Cell 46, 607-617 (1986)
STANDARD	full staff_review
REFERENCE	24 (sites; env gene conserved/variable regions; separate entries)
AUTHORS	Sternbach,B. R., Hahn,B. H., Shaw,G. M., McNeely,P. D., Modrow,S., Wolf,H., Parks,E. S., Parks,W. P., Josephs,S. F., Gallo,R. C. and Wong-Staal,F.
TITLE	Identification and characterization of conserved and variable

regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS
JOURNAL Cell 45, 637-648 (1986)
STANDARD full staff_review
REFERENCE 25 (sites; tar or transactivator target)
AUTHORS Wright,C. M. , Felber,B. K. , Paskalis,H. and Pavlakis,G. N.
TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus
JOURNAL Science 234, 998-992 (1986)
STANDARD full staff_review
REFERENCE 26 (sites; 3' orf mutations)
AUTHORS Terwilliger,E. , Sodroski,J. G. , Rosen,C. A. and Haseltine,W. A.
TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity
JOURNAL J. Virol. 60, 754-760 (1986)
STANDARD full staff_review
REFERENCE 27 (sites; pol p34 terminus)
AUTHORS Lightfoote,M. M. , Coligan,J. E. , Folks,T. M. , Fauci,A. S. , Martin,M. A. and Venkatesan,S.
TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus
JOURNAL J. Virol. 60, 771-775 (1986)
STANDARD full staff_review
REFERENCE 28 (sites; promoter; TAR, tat-III mutants)
AUTHORS Muesing,M. A. , Smith,D. H. and Capon,D. J.
TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein
JOURNAL Cell 48, 691-701 (1987)
STANDARD full staff_review
REFERENCE 29 (sites; envelope protein epitopes)
AUTHORS Modrow,S. , Hahn,B. H. , Shaw,G. M. , Gallo,R. C. , Wong-Staal,F. and Wall,F. H.
TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions
JOURNAL J. Virol. 61, 570-578 (1987)
STANDARD full staff_review
REFERENCE 30 (sites; trs/tat protein)
AUTHORS Goh,N. C. , Sodroski,J. G. , Rosen,C. A. and Haseltine,W. A.
TITLE Expression of the tat gene protein of human T-lymphotropic virus type (III) (HTLV-III/LAV) in bacteria
JOURNAL J. Virol. 61, 633-637 (1987)
STANDARD full staff_review
REFERENCE 31 (sites; inducible enhancer element)
AUTHORS Nabel,G. and Baltimore,D.
TITLE An inducible transcription factor activates expression of human immunodeficiency virus in T cells
JOURNAL Nature 326, 711-713 (1987)
STANDARD full staff_review
REFERENCE 32 (bases 5611 to 5611; revises [4])
AUTHORS Rather,L.
JOURNAL Unpublished (1987) Washington U Med School, St. Louis, MO
STANDARD full staff_review
REFERENCE 33 (sites; long terminal repeat)
AUTHORS Petarca,R. , Heath,C. , Goldenberg,G. J. , Rosen,C. A. , Sodroski,J. G. , Haseltine,W. A. and Hansen,U. M.
TITLE Transcription directed by the HIV long terminal repeat in vitro
AIDS Res. Hum. Retroviruses 3, 41-55 (1987)
STANDARD full staff_review
REFERENCE 34 (sites; R orf)
AUTHORS Wong-Staal,F. , Chanda,P. K. and Ghrayeb,J.
TITLE Human immunodeficiency virus: the eighth gene
AIDS Res. Hum. Retroviruses 3, 33-39 (1987)
STANDARD full staff_review

REFERENCE	35 (sites) sor
AUTHORS	Fisher,A. G. , Ensoli,B. , Ivanoff,L. , Chamberlain,M. , Petteway,S. , Ratner,L. , Gallo,R. C. and Wong-Staal,F.
TITLE	The sor gene of hiv-1 is required for efficient virus transmission in vitro
JOURNAL	Science 237, 888-893 (1987)
STANDARD	full staff_review
COMMENT	Sequence for [31 kindly provided in computer-readable form by L.Ratner, 19-AUG-1986. The HXB2 sequence is being used as a reference genome for all the HIV entries because it has been derived from a demonstrably infectious clone. Hence not all of the "sites" references above were concerned with this isolate.
FEATURES	from to/span description
pept	789 2291 gag polyprotein
pept	/ 2357 5095 pol polyprotein (NH2-terminus uncertain; AA at 2357)
pept	5040 5618 sor 23K protein
pept	5558 5794 R (ORF) protein
pept	5830 6044 tat protein, exon 2 (first expressed exon)
pept	8378 8423 tat protein, exon 3
pept	5980 6044 trs protein, exon 2 (first expressed exon)
pept	8378 8652 trs protein, exon 3
pept	6224 8794 envelope polyprotein
pept	8796 9167 27K protein (premature termination)
mRNA	455 9635 HXB2 genomic mRNA
pre-msg	455 9635 tat, trs, 27K subgenomic mRNA
IVS	6045 8377 tat intron 1
IVS	6045 8377 trs intron 2
IVS	6045 8377 27K mRNA intron 2
IVS	743 5776 tat,trs, 27K mRNA intron 1
IVS	6045 8377 tat, trs intron 2
LTR	1 634 5' LTR
LTR	5095 8718 3' LTR
rpt	454 551 R repeat 5' copy
rpt	9539 9635 R repeat 3' copy
binding	377 386 Sp1 binding site III
binding	388 397 Sp1 binding site II
binding	388 408 Sp1 binding site I
binding	636 653 primer (Lys-tRNA) binding site
revision	5611 5611 g in [32]; a in [4]
signal	5611 5616 HXB2 mRNA polyadenylation signal
BASE COUNT	3411 a 1773 c 2370 g 2164 t
ORIGIN	435 bp upstream of PvU1II site; 5' end of proviral genome.

Initial Score = 1858 Optimized Score = 2174 Significance = 0.00
Residue Identity = 88% Matches = 2243 Mismatches = 169
Gaps = 110 Conservative Substitutions = 0

X	10	20	30	40	50	60
AACAG-CAG---AAGACAGTGGCAATGAGAGTGAAGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA						
AATAAGACAGTTAATTGATAAGCTAATAGAAAAG-AGCAGAACAGTGGCAAT----GAGAGTGAAGGGAGAAA						
X 6180 6190 6200 6210 6220 6230 6240						

130 140 150 160 170 180
 CACAG-TCTATTATGGGTAC---CT---GTGTGCAA-----GGAAGCAA-CCACCA-CTCTATTTGTG
 ||||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TACAAAAAAATTCGTGCGTACAGTCTATTATGGGTACCTGTGGAAGGAAGCAACCACCACTCTATTTG
 6320 6330 6340 6350 6360 6370 6380

180 200 210 220 230 240
CATCAGATCTAAGCATA1GATA---CAGAGG-TACATA---AT--GTTTGGGCCACACATGCCTG--T
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
-----TGC---ATCAGATGCTAAAGCATATGATAACAGAGGTACATAATGTTGGC-CACA--CATGCCT
6380 6400 6410 6420 6430 6440

250 260 270 280 290 300 310
GTACCCACAGA-CCCCAACCAACAGAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGGAAAAA-
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GTGTACCCAUAGACCCCAACC-CA-CAAGAAGTA---GTA-TTG-----GTAAT-GTGACA---GAAAAT
6450 6460 6470 6480 6490

320 330 340 350 360 370
-ATGACATGCTAAACAGATG-CATGAGGATATAATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-T
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TTTACAT-GTGAA-AAATGACAT---GGTAGAA-CAGATGCATGAGGATATAA---TCAGTTTATGGGAT
6500 6510 6520 6530 6540 6550 6560

380 390 400 410 420 430 440
AAAA---TAAACCCCACCTCTGTGTTAGTTAA-AGTGCAC TGATTTGG---GGAATGCTACTAAT---ACCAA
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
CAAAGCCTAAACCA---TGTGAAAATTAACCCCACCTCTGTGTTAGTTAAAGTGC-ACTGATTGAAGAA
6570 6580 6590 6600 6610 6620 6630

450 460 470 480 490 500 510
TACTAGTAATACCAATAGTAGTAGTACGGGGAAATGATGATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCAA
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TGATACATAACCAATAGTAGTAAGGGAGAAATGATAATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCAA
6640 6650 6660 6670 6680 6690 6700

520 530 540 550 560 570 580
TATCAGCACAGNATAYAGAGTAAAGGTGCAGAAAGAATATGCATTTTTTATAAAACTTGATATAATACCAAT
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TATCAGCACAGCATAAGAGCTAAGGTGCAGAAAGAATATGCATTTTTTATAAAACTTGATATAATACCAAT
6710 6720 6730 6740 6750 6760 6770

590 600 610 620 630 640 650
AGATAATGATACTACCAAGCTATACTGTTGACAACTTGTAAACACCTCAGTCATTACACAGGGCCTGTCCAAAGGT
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
AGATAATGATACTACCAAGCTATACTGTTGACAAAGTTGTAAACACCTCAGTCATTACACAGGGCCTGTCCAAAGGT
6780 6790 6800 6810 6820 6830 6840

660 670 680 690 700 710 720 730
ATCCTTGGCCAAATCCCATACATTATTGTGCCCCGGCTGGTTTGCATTCTAAATGTAATAATAAGAC
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
ATCCTTGGCCAAATCCCATACATTATTGTGCCCCGGCTGGTTTGCATTCTAAATGTAATAATAAGAC
6850 6860 6870 6880 6890 6900 6910

740 750 760 770 780 790 800
GTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATC
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATC
6920 6930 6940 6950 6960 6970 6980 6990

810 820 830 840 850 860 870
AACTCAACTGCTTTGAAATGTCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAA
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
AACTCAACGCTGTTAYATGTCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGTCAATTTCACGGACAA
7000 7010 7020 7030 7040 7050 7060

880 890 900 910 920 930 940
TCTAAACCATATAAGTACAGCTGAAACCAATCTGTAGAAATTAAATTGTACAAGACCCAACAACAATAAG
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TCTAAACCATATAAGTACAGCTGAAACCAATCTGTAGAAATTAAATTGTACAAGACCCAACAACAATAAG
7070 7080 7090 7100 7110 7120 7130

950 960 970 980 990 1000 1010
AAAAAGTAATCCGATACCAAGAGGGGACCGAGGAGCATTGTTACAATAGGAAAAATAGGAAATATGAGACA
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAAAAGAATCCGATACCAAGAGGGGACCGAGGAGCATTGTTACAATAGGAAAAATAGGAAATATGAGACA
7140 7150 7160 7170 7180 7190 7200

1020 1030 1040 1050 1060 1070 1080 1090
AGCACATTCTAACATTAACTAGCAAAATCAATGCCACTTAAAACAGATAGCTAGCAAATTAAGAGAAC
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AGCACATTCTAACATTAACTAGCAAAATCAACACTTAAAACAGATAGCTAGCAAATTAAGAGAAC
7210 7220 7230 7240 7250 7260 7270

1100 1110 1120 1130 1140 1150 1160
ATTTGGAAATAATAACAATACTTAAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTT
||||||| ||||| ||||| ||||| ||||| ||||| |||||
ATTCGGAAATAATAACAATACTTAAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTT
7280 7290 7300 7310 7320 7330 7340 7350

1170 1180 1190 1200 1210 1220 1230
TAATTGTGAGGGAAATTTCTACTGTAATTCAACACAACGTGTTAATAGTACTTGGTTAATAGTACTTG
||||||| ||||| ||||| ||||| ||||| ||||| |||||
TAATTGTGAGGGAAATTTCTACTGTAATTCAACACAACGTGTTAATAGTACTTGGTTAATAGTACTTG
7360 7370 7380 7390 7400 7410 7420

1240 1250 1260 1270 1280 1290 1300
GAG1ACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCAGCAGAATAAAACAATT
||||||| ||||| ||||| ||||| ||||| ||||| |||||
GAG1ACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCCTCCCAGCAGAATAAAACAATT
7430 7440 7450 7460 7470 7480 7490

1310 1320 1330 1340 1350 1360 1370
AAACATGTGAGGGAAAGTAGGAAAAACAAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCATCAA
||||||| ||||| ||||| ||||| ||||| ||||| |||||
AAACATGTGAGGGAAAGTAGGAAAAACAAATGTATGCCCTCCCATCAGTGGACAAATTAGATGTTCATCAA
7500 7510 7520 7530 7540 7550 7560

1380 1390 1400 1410 1420 1430 1440 1450
TATTACAGGGCTGCTATTAAACAGAGATGGTGGTAATAACACAAATGGTCCGAGATCTCAGACCTGGAGG
||||||| ||||| ||||| ||||| ||||| ||||| |||||
TATTACAGGGCTGCTATTAAACAGAGATGGTGGTAATAACACAAATGAGTCCGAGATCTCAGACTGGAGG
7570 7580 7590 7600 7610 7620 7630

1460 1470 1480 1490 1500 1510 1520
AGGAAGATAAGGGGACAAATTGAGAAGTGAAATTATATAAAGTAGTAAAAATTGAAACCTAGGAGT
||||||| ||||| ||||| ||||| ||||| ||||| |||||
AGGAAGATAAGGGGACAAATTGAGAAGTGAAATTATATAAAGTAGTAAAAATTGAAACCTAGGAGT
7640 7650 7660 7670 7680 7690 7700 7710

1530 1540 1550 1560 1570 1580 1590
AGCACCCACCAAGGCAAGAGAGAGAGTGCTGCGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTC
||||||| ||||| ||||| ||||| ||||| ||||| |||||
AGCACCCACCAAGGCAAGAGAGAGAGTGCTGCGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTC
7720 7730 7740 7750 7760 7770 7780

1600 1610 1620 1630 1640 1650 1660
TGGGTTCTTGCGACGAGGAAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAAATT
||||||| ||||| ||||| ||||| ||||| ||||| |||||
TGGGTTCTTGCGACGAGGAAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAAATT
7790 7800 7810 7820 7830 7840 7850

1670 1680 1690 1700 1710 1720 1730
ATTGTCGAGGTTAGTGCGAGGAGGAAACAAATTGCTGAGGGCTATTGAGGCGAACAGCATTGTTGCAACT
||||||| ||||| ||||| ||||| ||||| ||||| |||||
ATTGTCGAGGTTAGTGCGAGGAGGAAACAAATTGCTGAGGGCTATTGAGGCGAACAGCATTGTTGCAACT
7860 7870 7880 7890 7900 7910 7920

9. KUNZ-158-CL32. SEQ

HIVSC Human immunodeficiency virus type 1, isolate SC (3'

LOCUS HIVSC 4273 bp ss-RNA **VRL** 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate SC (3' end of genome).
ACCESSION M17450
KEYWORDS.
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate SC, provial DNA.
ORGANISM Human immunodeficiency virus type 1
VIRIDAE; ss-RNA enveloped viruses; Retroviridae;
Lentivirinae.
REFERENCE 1 (bases 1 to 4273)
AUTHORS Gungo,C., Guo,H.-B., Franchini,G., Aldovini,A., Collalti,E., Farrell,K., Wong-Staal,F., Gallo,R.C. and Reitz,M.S.Jr.
TITLE Envelope sequences of two new United States HIV-1 isolates
JOURNAL Virology 164, 531-536 (1988)
STANDARD full staff_review
COMMENT Kindly made available in computer readable form by Marv Reitz,
N. C. I., Bethesda, MD 20892 U. S. A. This isolate was taken from a
California AIDS patient in 1984. There is an in-frame stop codon at
position 3212 of the envelope coding sequence; the nef cds is
uncertain beyond position 4049. A stop codon, 'taa,' in-frame with
the nef sequence does exist at positions 4224-4226.
FEATURES

	from	to/span	description					
pept	<	1	330	vif protein (partial; AA at 1)				
pept		270	560	vpr protein				
pept		541	755	tat protein, exon 2 (first expressed exon)				
		3088	3173	tat protein, exon 3 (AA at 3090)				
pept		680	755	rev protein, exon 2 (first expressed exon)				
		3088	3303	rev protein, exon 3 (AA at 3091)				
pept		772	864	vpu protein (premature termination)				
pept_ps		835	3505	envelope polyprotein (premature stop at 3212)				
pept		3507	4226	nef protein				
pre-msg	<	1	> 4273	genomic mRNA				
pre-msg	<	1	> 4273	tat, rev, nef subgenomic mRNA				
IVS	<	1	497	tat, rev, nef subgenomic mRNA intron 1				
IVS		756	3088	tat cds intron 2				
IVS		756	3088	rev cds intron 2				
IVS		758	3088	tat, rev, nef subgenomic mRNA intron 2				
LTR	3786	>	4273	3' LTR				
rpt	4248	>	4273	R repeat 3' copy				
site	3212		3214	premature stop (tag) in env cds				
BASE COUNT	1447	B	760	C	1053	G	1013	T

ORIGIN

Initial Score = 1158 Optimized Score = 2151 Significance = 0.00
Residue Identity = 88% Matches = 2201 Mismatches = 214
Gaps = 82 Conservative Substitutions = 0

X	10	20	30	40	50	60
AAGAAGCAG-AAGA-CAGTGGCAATGAGAGTGAAGGAGAA-ATATC----AGCACCTGTGGAGATGGGGGTGG						
AAT>AGAGAAAGAGCAGAGAACATTGGCAATGAGAGTGAAGGGATCAGGAAGGAATTAT-CAG-CACTTGTGG						
810	820	830	840	850	860	870
70	80	90	100	110	120	130
AAA1GGGGCACCACTGCTCCTTGGEATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTAT						
AGATGGGGCAAGATGCTCCTGGGGATATTGATGATCTGTAGTGCTGCAGAACAAATTGTGGGTACAGTCTAT						
880	890	900	910	920	930	940
140	150	160	170	180	190	200
TATTTGGGATGTCCTTGGAAGGGAAACCAACACTCTATTTGATCATCAGATGCTAAAGCATATGATA						

TATGGGTACCTGTTGAAAGAACCAACCACTCTATTGGTGCATCAGATGCTAAGCATATGATA
1050 1060 1070 1080 1090 1100 1110 1120

210 220 230 240 250 260 270 280
GAGGTACATTAATGTTGGGCCACACATGCTGTACCCACAGACCCAAACCCACAAGAAGTAGTATTGGTA
1130 1140 1150 1160 1170 1180 1190

290 300 310 320 330 340 350
AATGTGACAGAAAATTTAACATGTTGGAAAAATGACATGGTAGAACAGATGCATGAGGAATAATCAGTTA
1200 1210 1220 1230 1240 1250 1260

360 370 380 390 400 410 420
TGGGNTCAAGGCTAAAGCCATGTTAAATTAAACCCCACACTCTGTGTTAGTTAAAGTGCACTGATTGGGG
1270 1280 1290 1300 1310 1320 1330

430 440 450 460 470 480 490
AATGCTACTAATACCAATACTATAATACCAATAGTAGT-AGCGGGGAAATGATGATGGAGAAAGGGAGAGAT
1340 1350 1360 1370 1380 1390 1400

500 510 520 530 540 550 560
AAAAAACTCTCTTTCAATAATCAGCACAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTATAA
1410 1420 1430 1440 1450 1460 1470

570 580 590 600 610 620 630 640
ACTTGATATAATACCAATAGATAATGATACTACCAGCTACGTTGACAAGTTGTAACACCTCAGTCATTAC
1480 1490 1500 1510 1520 1530 1540

650 660 670 680 690 700 710
ACAGAGCTATCCTTAAAGTATCCTTGGCCAAATTCCCATACATTATTGTGCCCGGCTGGTTTGCATTCT
1550 1560 1570 1580 1590 1600 1610

720 730 740 750 760 770 780
AAATGTAAATAATACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGG
1620 1630 1640 1650 1660 1670 1680

790 800 810 820 830 840 850
AATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAACAGAGGTAGTAATTAGATC
1690 1700 1710 1720 1730 1740 1750

860 870 880 890 900 910 920
TCCCAATTTCAGAACATGCTAAACATAATAGTACAGCTGAACCAATCTGTAGAAATTAAATTGTACAAG
1760 1770 1780 1790 1800 1810 1820

930 940 950 960 970 980 990
TGGGAAATAACGAAACAGGACATTTGTAATCAGAACGGGACCAAGGGAGAGCATTTGTACAATAGGA--

1600 1610 1620 1630 1640 1650 1660 1670 1680 1690
 GCGGAAACAAAGAATACACCAAGAAGTATAUATAT--AG---GACCAGGGAGAGCATTATGCAACAGGAGA
 1630 1640 1650 1660 1670 1680 1690
 1000 1010 1020 1030 1040 1050 1060
 -AATATAGGAAATATAGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTAAACAGAT
 1010 1020 1030 1040 1050 1060
 CATATAGGAAATATAAGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTAAACAGAT
 1030 1040 1050 1060
 1070 1080 1090 1100 1110 1120 1130 1140
 AGCTAGCAATTAAAGAGAACATTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCC
 1110 1120 1130 1140
 AGTATAAAATTAAGAGAACATTG---AGAATAAAACAATAATCTTTAATCGATCCTCAGGAGGGAGACCC
 1090 1100 1110 1120 1130 1140
 1150 1160 1170 1180 1190 1200 1210
 AGAAATTETAAUGCACAGTTTAATTGTGGAGGGAAATTTTCTACTGTAATTCAACACAACGTGTTAATAG
 1160 1170 1180 1190 1200 1210
 AGAAATTGTAATTCACAGTTTAATTGTGGAGGGAAATTTTCTACTGTAATTCAACACAACGTGTTAGTAG
 2040 2050 2060 2070 2080 2090 2100
 1220 1230 1240 1250 1260
 TACTTGGTCTAACTGACT---TGG---AGTA---CTGAAGG---GTCA-AATAACAC---TGAAG---G
 2110 2120 2130 2140 2150 2160 2170
 TACTTGG---AACTGACTGAGGGTCATAACACTGGAGGGAAATGACACAAATCACCCTCCCATGCGAAAT
 1270 1280 1290 1300 1310 1320 1330
 AAATGACACAAAT-CACA-CTCCCACTCAGAGAAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAAAAGCA
 1310 1320 1330
 AAATGAAATTATTAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAA--AGGACAAG--
 2180 2190 2200 2210 2220 2230 2240
 1340 1350 1360 1370 1380 1390 1400
 ATGATGCGCTCCATCAGCGAAATAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAAGAGAT
 2250 2260 2270 2280 2290 2300
 TTAATGTTCATCAAAT---ATTACA---GGGCTGCT-AT-TAACAAAGAGATGG-TGGTAATAGCAAGA-AT
 1410 1420 1430 1440 1450 1460 1470
 GGCGTGAAATACACAAATGGCTCCAGAGATCTTCAGACCTGGAGGGAGGAGATATGAGGGACAATTGGAGAAGT
 2310 2320 2330 2340 2350 2360 2370
 GGCGAAATGAGAGA---CACAGAGATCTTAGACCGGGAGGGAGGAGATATGAGGGACAATTGGAGAAGT
 1480 1490 1500 1510 1520 1530 1540 1550
 GAATTATAATAATATAAGTAGTAAATAATTGAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTG
 2380 2390 2400 2410 2420 2430 2440
 GAATTATAATAATATAAGTAGTAAATAATTGAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTG
 1560 1570 1580 1590 1600 1610
 GTGGAG
 2450 2460 2470 2480 2490 2500 2510
 GTGGAG
 1620 1630 1640 1650 1660 1670 1680 1690
 ACTATGGCGCAGGGTCAATGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAG
 1630 1640 1650 1660 1670 1680 1690
 ACTATGGCGCAGGGTCAATGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAG
 2520 2530 2540 2550 2560 2570 2580 2590
 1700 1710 1720 1730 1740 1750 1760
 AAGCGATTTCTTACGGTACAGGCCAGACAATTGTCTGGTATAGTGCAGCAGCAG

AACAAATTCTGAGGACTATTGAGGGCGCAACACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTC
 2600 2610 2630 2630 2640 2650 2660
 1770 1780 1790 1800 1810 1820 1830
 CAGGCAAGGAAATCCCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGA
 2670 2690 2690 2700 2710 2720 2730
 1840 1850 1860 1870 1880 1890 1900
 ARAACTCATTTGGGCGCACTGCTGGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTGGAAAT
 2740 2750 2760 2770 2780 2790 2800
 1910 1920 1930 1940 1950 1960 1970
 AACATGACCTGAACTGGAGTGCGACAGAAAATTAAACAATTACACAAGCTTAATACATTCTTAATTGAAGAA
 2810 2820 2830 2840 2850 2860 2870
 1980 1990 2000 2010 2020 2030 2040 2050
 TGGAAAAACCAAAAGAAATGAAACAAGAATTGGAAATTAGATAAAATGGCAAGTTGTGGAAATTGG
 2880 2890 2900 2910 2920 2930 2940 2950
 2060 2070 2080 2090 2100 2110 2120
 TTAAACATAACAAATTGGCTTGTGATAATTGATAATGATAGTAGGAGGGCTGGTAGGTTAAAGA
 2960 2970 2980 2990 3000 3010 3020
 2130 2140 2150 2160 2170 2180 2190
 ATATTTTCTGCTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCCAC
 3030 3040 3050 3060 3070 3080 3090
 2200 2210 2220 2230 2240 2250 2260
 CTCCCCAACCTGGAGGCGACAGGGCCCGAAGGAATAGAAGAAGGAGGAGAGACAGAGACAGA
 3100 3110 3120 3130 3140 3150 3160
 2270 2280 2290 2300 2310 2320 2330
 TCCATTCTTAACTGACGAACTGGCACTTGTGGACGATCTGGAGGCCTTGTGCCTCTCAGCTA
 3170 3180 3190 3200 3210 3220 3230
 2340 2350 2360 2370 2380 2390 2400 2410
 CCACUGCGCTGGAGAGACTTACTCTTGTAAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGC
 3240 3250 3260 3270 3280 3290 3300
 2420 2430 2440 2450 2460
 CCTGAAATATTGAGGAACTCTACAGTATTGGAGTCAGGAACTAAGG
 3310 3320 3330 3340 3350 X

HIVMAL Human immunodeficiency virus type 1, isolate MAL,

LOCUS HIVMAL 8229 bp ss-RNA **VRL** 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate MAL, complete genome.
ACCESSION K03436
KEYWORDS
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate MAL, proviral DNA, clone M-H11.
ORGANISM Human immunodeficiency virus type 1
MATERIALS ss-RNA enveloped viruses; Retroviridae;
Lentivirinae.
REFERENCE 1 (bases 1 to 8229)
AUTHORS Allizon, M., Wain-Hobson, S., Montagnier, L. and Sonigo, P.
TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients
JOURNAL Cell 45: 63-74 (1986)
STANDARD full staff_review
COMMENT A region similar to the vpu cds of HIVELI exists from positions 5838 (starts with 'ata' instead of 'atg') to 5881. To date it has not been possible to construct an infectious clone for the MAL viral genome (Keith Peden, Johns Hopkins University School of Medicine, Baltimore, MD).
FEATURES

	from	to/span	description
pept	350	1367	gag polyprotein
pept	< 1663	4671	pol polyprotein (NH2-terminus uncertain; AA at 1663)
pept	4618	5184	vif protein
pept	5134	5424	vpr protein
pept	5405	5519	tat protein, exon 2 (first expressed exon)
	7558	8007	tat protein, exon 3 (AA at 7960)
pept	5544	5619	rev protein, exon 2 (first expressed exon)
	7958	8236	rev protein, exon 3 (AA at 7961)
pept	< 5636	5681	vpu protein (AA at 5636)
pept	5799	8378	envelope polyprotein precursor
pept	8380	8909	nef protein
pre-msg	1	3228	genomic mRNA
pre-msg	1	3228	tat, rev, nef subgenomic mRNA
IVS	308	5351	tat, rev, nef subgenomic mRNA intron 1
IVS	5820	7956	tat cds intron 2
IVS	5820	7958	rev cds intron 2
IVS	5820	7959	tat, rev, nef subgenomic mRNA intron 2
LTR	< 1	177	5' LTR
LTR	8576	> 5229	3' LTR
rpt	< 1	35	R repeat 5' copy
rpt	9133	9229	R repeat 3' copy
binding	179	186	primer (Lys-tRNA) binding site
signal	9204	9208	mRNA polyadenylation signal
BASE COUNT	3355 a	1627 c	2204 g 2043 t
ORIGIN	Cap site of genomic RNA.		

Initial Score = 916 Optimized Score = 2060 Significance = 0.00
Residue Identity = 84% Matches = 2085 Mismatches = 350
Gaps = 43 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
AAGAGCAGAAGACAGTGGCAATGAGAGTGAGGAGAAAATATCAGCACCTGTGGAGATGGGGGTGGAAATGGG							
TT							
AAAGAGCAGAAGACAGTGGCAATGAGAGTGAGGAGAGATACA-GAGGAATTATCAAAA--CTGGTGAGATGGG							
5780 5790 5800 5810 5820 5830 5840							
80	90	100	110	120	130	140	
GGCGCATGCTTGGAGATACTGATGATCTGTAGTGCTACAGAAAAATTGTGGGGTCACAGTCTATTATGGGG							
TT							
GCAATGATGCTTGGAGATTTGACTGTAGTATTGCAGAAGATTTGTGGGGTACAGTTATTATGGGG							
5850 5860 5870 5880 5890 5900 5910							

560	560	560	570	580	590	
GGCGGAAACAGAATACAGAGAAAAGCTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTGTTACAATAGGAA						
6710	6720	6730	6740	6750	6760	6770
1000	1010	1020	1030	1040	1050	1060
AAATAGGAAAHATGAGAACAAAGCATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTAAACAGATAG						
6780	6790	6800	6810	6820	6830	6840
1080	1090	1100	1110	1120	1130	1140
CTACGAAAATTAAGAGAACATTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAG						
6850	6860	6870	6880	6890	6900	6910
1150	1160	1170	1180	1190	1200	1210
AAATTGTAAACGACAGTTTAAATTGTGGAGGGAAATTTTCTACTGTAATTCAACACAACGTGTTAATAGTA						
6920	6930	6940	6950	6960	6970	6980
1220	1230	1240	1250	1260	1270	1280
CTTGCGTTAATAGTACTTGGAGTACTGAAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCAT						
6990	7000	7010	7020	7030	7040	7050
1290	1300	1310	1320	1330	1340	1350
GGCGGAAACAGAATTTATAAACATGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGAC						
7060	7070	7080	7090	7100	7110	7120
1360	1370	1380	1390	1400	1410	1420
AAATTAGATTTGTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACA---ACAAT-GGG						
7130	7140	7150	7160	7170	7180	7190
1430	1440	1450	1460	1470	1480	1490
TC---CGAGATCTTCAGACCTGGAGGGAGGAGATATGAGGGACAATTGGAGAAGTGAAATTATATAAATATAAAG						
7200	7210	7220	7230	7240	7250	7260
1500	1510	1520	1530	1540	1550	1560
TAGTAAAAAATTGAGCATTAGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGTTGCAGAGAGAAAAAGAG						
7270	7280	7290	7300	7310	7320	7330
1570	1580	1590	1600	1610	1620	1630
CACTGGGAATTAACGCTTTGTTCCCTTGCGGTCTGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGA						
7340	7350	7360	7370	7380	7390	7400
1650	1660	1670	1680	1690	1700	1710
CGCTTACCGTCAAGGCAAGACAATTATTGTCTGGTAGTACTGTCAGCAGCAGAACATTTGCTGAGGGCTATTG						
7420	7430	7440	7450	7460	7470	7480

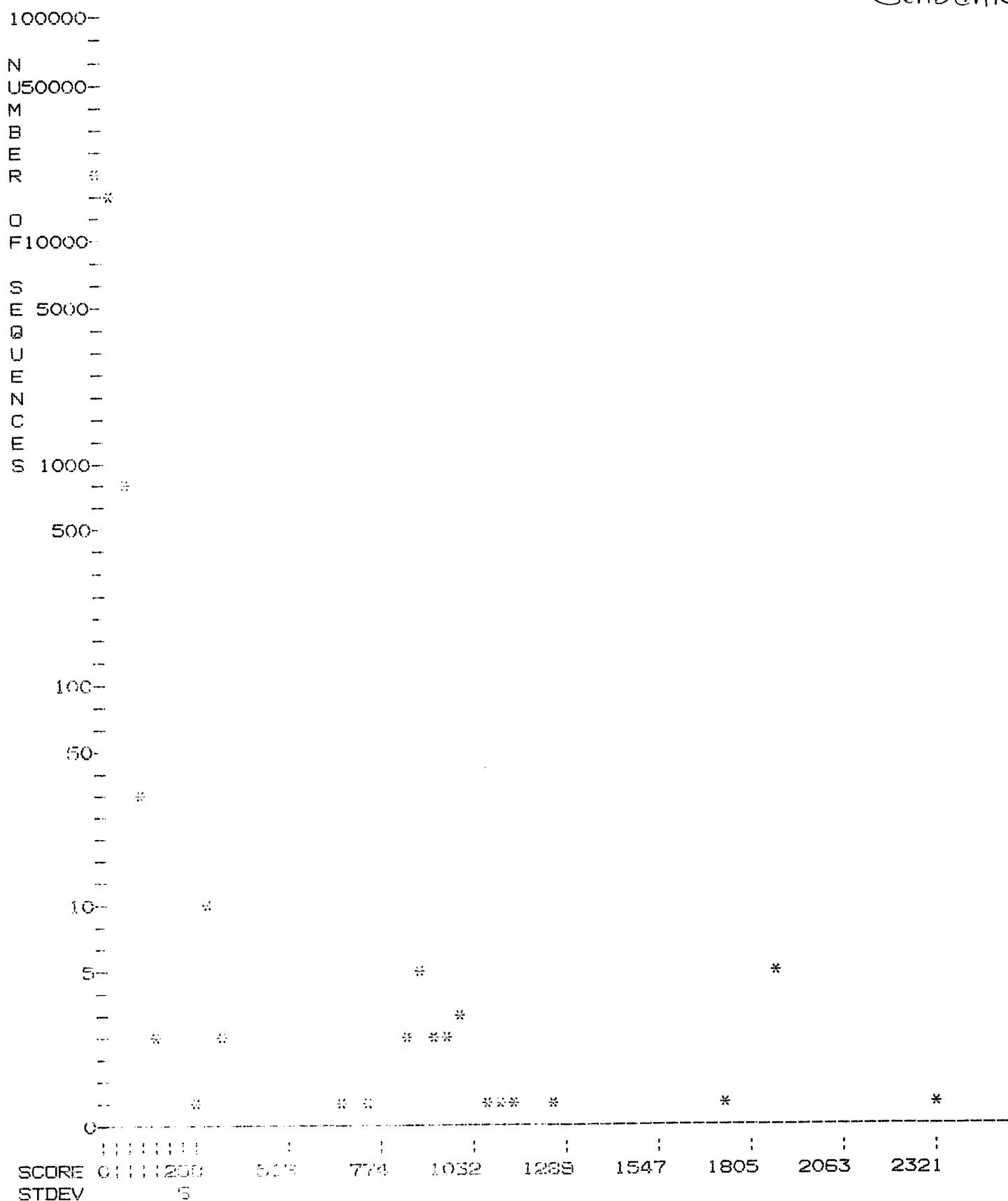
1720 1730 1740 1750 1760 1770 1780
 AGGCGCAACGAGCTCTTGTAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGG
 1790 1800 1810 1820 1830 1840 1850
 AGGUGCAACGAGCTCTTGTAACTCACGGTCTGGGGCATTAACAGCTCCAGGCAAGAGTCCTGGCTGTGG
 1860 1870 1880 1890 1900 1910 1920
 AAABATACUTTAAGGATCAACGAGCTGGGATTGGGGTTGCTCTGGAAAACACTCATTTGCACCACTGCTG
 1930 1940 1950 1960 1970 1980 1990 2000
 TGCUTTGCGATTTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGAATAACATGACCTGGATGGAGTGGG
 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070
 TGCUTTGCGATTTAGTTGGAGTAATAAAATCTCTAGATGACATTGGAATAATATGACCTGGATGCAGTGGG
 2030 2040 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140
 ACAGAGAAATTAGCAATTACACAAAGCTTAAATACATTCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGA
 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210
 AAAAGAAATTAGCAATTACACAGCATAATATACTTAATTGAAGAATCGCAAATCCAGCAAGAAAAGA
 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280
 2230 2240 2250 2260 2270 2280 2290 2300 2310 2320 2330 2340 2350
 ATGCGATTTAGCTTACGAGCTTACGAGCTTACGAGCTTACGAGCTTACGAGCTTACGAGCTTACGAGCTTACG
 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400 2410 2420 2430
 CGACAGGCGGGAGGAAATAAGAAGGAG
 2380 2390 2400 2410 2420 2430 2440 2450 X
 CTCTTAACTGCTTACGAGGATTGTAACCTCTGGGGACGATCTGGGGAGCCTTGTGCCTCTTCAGCTACCCACCGCTTGTGAGAGACTTA
 2440 2450 X
 CTCTTAACTGCTTACGAGGATTGTAACCTCTGGGGACGATCTGGGGAGCCTTGTGAGAGACTTA
 2450 X

Query sequence being compared: KUNZ-158-CL33. SEQ
 Number of sequences searched: 31228
 Number of entries above cutoff: 38

Kunz
07/158652

Results of the initial comparison of KUNZ-158-CL33. SEQ with: Claim 33
 Data bank: GenBank 82.0; all entries

Genbank



PARAMETERS

Similarity matrix	Unify	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.43		
Cutoff score	146		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	37	36	19.76

Times:	CPU	Total Elapsed
	00:53:20.08	02:16:33.00

Number of molecules:	37183850
Number of sequences searched:	31228
Number of entries above cutoff:	38

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.					
		Length	Score	Score	Sig.	Frame	
1. H1VBRUCE	**** 115 standard deviations above mean ***	9229	2321	2436	115.50	0	
2. H1VPV22	**** 93 standard deviations above mean ***	9770	1877	2180	93.04	0	
3. H1VH089	**** 92 standard deviations above mean ***	3156	1873	2176	92.84	0	
4. HIVSH102	Human immunodeficiency virus t	8932	1872	2176	92.79	0	
5. HIVKRE05	Human immunodeficiency virus t	9718	1858	2164	92.08	0	
6. HIVNL43	**** 85 standard deviations above mean ***	9709	1729	2169	85.56	0	
7. HTVELICG	**** 61 standard deviations above mean ***	9176	1246	1893	61.14	0	
8. HTVSC	**** 58 standard deviations above mean ***	4273	1159	2139	56.74	0	
9. HIVH3PH8	**** 54 standard deviations above mean ***	3563	1112	1761	54.36	0	
10. HIVZ371	**** 52 standard deviations above mean ***	3457	1066	1959	52.03	0	
11. HIVMN09	**** 48 standard deviations above mean ***	9738	996	2203	48.49	0	
12. HIVZ2ZC	**** 47 standard deviations above mean ***	9081	983	1908	47.84	0	
13. HIVZ6	Human immunodeficiency virus t	5159	983	1915	47.84	0	
14. HIVJH32	Human immunodeficiency virus t	2903	975	1365	47.43	0	
15. HTVJY1	**** 45 standard deviations above mean ***	2653	972	1917	47.28	0	
16. HIVDD012	**** 44 standard deviations above mean ***	3373	932	2191	45.26	0	
17. HIVM9L	**** 43 standard deviations above mean ***	9229	916	2041	44.45	0	

18. HIVRFENV	Human immunodeficiency virus type 1	2622	887	1567	42.98	O
19. HIVRF	Human immunodeficiency virus type 1	9128	887	1567	42.98	O
20. HIVSF20G	Human immunodeficiency virus type 1	9737	883	1941	42.78	O

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init.	Opt.	Sig.	Frame
			Score	Score		
1. HIVBRUG	Human immunodeficiency virus type 1	9229	2321	2436	0.00	O
2. HIVMN09	Human immunodeficiency virus type 1	9738	996	2203	0.00	O
3. HTVCDG4E	Human immunodeficiency virus type 1	3373	932	2191	0.00	O
4. HIVPV22	Human immunodeficiency virus type 1	9770	1877	2180	0.00	O
5. HIVBH102	Human immunodeficiency virus type 1	8932	1872	2176	0.00	O
6. HIVHXB3	Human immunodeficiency virus type 1	3156	1873	2176	0.00	O
7. HIVNL43	Human immunodeficiency virus type 1	9709	1729	2169	0.00	O
8. HIVHXBRC6	Human immunodeficiency virus type 1	9718	1858	2164	0.00	O
9. HIV50	Human immunodeficiency virus type 1	4273	1159	2139	0.00	O
10. HIVMA1	Human immunodeficiency virus type 1	9229	916	2041	0.00	O
11. HIVBRVA	Human immunodeficiency virus type 1	3600	861	2027	0.00	O
12. HIVZ321	Human immunodeficiency virus type 1	3457	1066	1959	0.00	O
13. HIVSF20G	Human immunodeficiency virus type 1	9737	883	1941	0.00	O
14. HIVJY1	Human immunodeficiency virus type 1	2653	972	1917	0.00	O
15. HIVZE	Human immunodeficiency virus type 1	5159	983	1915	0.00	O
16. HIVZZR	Human immunodeficiency virus type 1	9081	983	1908	0.00	O
17. HIVELING	Human immunodeficiency virus type 1	9176	1246	1893	0.00	O
18. HIVH3BHQ	Human immunodeficiency virus type 1	3563	1112	1761	0.00	O
19. HIVRFENV	Human immunodeficiency virus type 1	2622	887	1567	0.00	O
20. HIVRF	Human immunodeficiency virus type 1	9128	887	1567	0.00	O

1. KUNZ-158-CL33, SE12

HIVBRUG Human immunodeficiency virus type 1, isolate BRU,

LOCUS	HIVBRUGR	9229 bp ss-RNA	VRL	15-JUN-1989
DEFINITION		Human immunodeficiency virus type 1, isolate BRU, complete genome (LAV-1).		
ACCESSION	KO2013			
KEYWORDS		TAP protein; TAT protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activator.		
SOURCE		Human immunodeficiency virus type 1 (HIV-1), isolate BRU (LAV-1), proviral DNA clone lambda-J19.		
ORGANISM		Human immunodeficiency virus type 1 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae		
REFERENCE		[1] (bases 1 to 9229)		
AUTHORS		Vain-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M.		
TITLE		Nucleotide sequence of the AIDS Virus, LAV		
JOURNAL		Cell 40, 9-17 (1985)		
STANDARD		full staff_review		
REFERENCE		[2] (bases 1712 to 1749; revision of [1])		
AUTHORS		Alizon,M., Vain-Hobson,S., Montagnier,L. and Sonigo,P.		
TITLE		Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients		
JOURNAL		Cell 46, 63-74 (1986)		
STANDARD		full staff_review		
COMMENT		The original LAV, sometimes called LAV-1 to distinguish it from HIV-1 (LAV-2), is now referred to as HIV-1bru. An infectious clone		

of this virus has been constructed by Keith Peden, Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205 (301) 955-3652. HIVNL43 is also an infectious clone having for its 3' half a clone of the BRU isolate. Acquired immune deficiency syndrome (AIDS) is caused by a retrovirus known by several different names, probably representing two separate strains: human T-cell lymphotropic virus-III (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the other. All three viruses, whose sequences do not differ by more than about 6%, are believed to belong to the retroviral subfamily Lentiviridae, or "slow" viruses.

For the details of the annotation and for other pertinent references, see the HIV reference entry.

FEATURES	from	to/span	description
pept	336	1874	gag polyprotein
pept	< 1631	4578	pol polyprotein (NH2-terminus uncertain; AA at 1631)
pept	4623	5201	vif protein
pept	5141	5431	vpr protein
pept	5412	5626	tat protein, exon 2 (first expressed exon)
	7972	8017	tat protein, exon 3 (AA at 7973)
pept	5551	5626	rev protein, exon 2 (first expressed exon)
	7972	8246	rev protein, exon 3 (AA at 7974)
pept	5643	5696	vpu protein
pept	5803	6380	envelope polyprotein
pept	6390	8010	nef protein
pre-msg	1	9229	genomic mRNA
pre-msg	1	9229	tat, rev, nef subgenomic mRNA
IVS	290	5350	tat, rev, nef subgenomic mRNA intron 1
IVS	5627	7971	tat cds intron 2
IVS	5627	7571	rev cds intron 2
IVS	5627	7571	tat, rev, nef subgenomic mRNA intron 2
LTR	< 1	180	5' LTR
LTR	8679	> 9223	3' LTR
rpt	< 1	87	R repeat 5' copy
rpt	9133	9223	R repeat 3' copy
binding	182	199	primer (Lys-tRNA) binding site
revision	1712	1743	atttcttcagagcagaccagaccaacagccccccacag in [2]; ag in [1]
signal	8205	8210	mRNA polyadenylation signal
BASE COUNT	3269 a	1658 c	2232 g 2052 t
ORIGIN	Cap site of genomic RNA.		

Initial Score = 2321 Optimized Score = 2436 Significance = 0.00
 Residue Identity = 99% Matches = 2436 Mismatches = 3
 Gaps = 1 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
ATGAGAGTCAGAGAGAAATATCAACACTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA							
ATGAGAGTCAGAGAGAAATATCAACACTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA							
X 5810 5820 5830 5840 5850 5860 5870							
80	80	100	110	120	130	140	
TTGAATCTGTAGTCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGGAAGCA							
TTEATGATCTGTAGTCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGGAAGCA							
5880 5890 5900 5910 5920 5930 5940							
150	160	170	180	190	200	210	
ACCAACACTCTATTTATGCATCAAGATCTAAAGCATATGATACAGAGGTACATAATGTTGGGCCACACAT							
ACCAACACTCTATTTATGCATCAAGATCTAAAGCATATGATACAGAGGTACATAATGTTGGGCCACACAT							
5950 5960 5970 5980 5990 6000 6010							

220 230 240 250 260 270 280
GCCTGTACCCACAGACCCAAACCAACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGG
|||||
GCCTGTACCCACAGACCCAAACCAACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGG
6020 6030 6040 6050 6060 6070 6080 6090
290 300 310 320 330 340 350 360
AAAATGACATGATGATGATGACAGATGATGAGGATATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
|||||
AAAATGACATGATGATGACAGATGATGAGGATATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
6100 6110 6120 6130 6140 6150 6160
370 380 390 400 410 420 430
AAATTAACCCACTCTGTAGT/TAAGTGCACTGATTGGGAATGCTACTAATACCAATACTAGTAAT
|||||
AAATTAACCCACTCTGTAGT/TAAGTGCACTGATTGGGAATGCTACTAATACCAATACTAGTAAT
6170 6180 6190 6200 6210 6220 6230
440 450 460 470 480 490 500
ACCAATGATGATGCGGGAAATGATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCATAATATCAGCACA
|||||
ACCAATGATGATGCGGGAAATGATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCATAATATCAGCACA
6240 6250 6260 6270 6280 6290 6300
510 520 530 540 550 560 570
AGNATAAGAGGTAAAGTGCAAAAGAAATATGCATTTTTATAAACTTGTATATAATACCAATAGATAATGAT
|||
AGCATAGAGGTAAAGTGCAAAAGAAATATGCATTTTTATAAACTTGTATATAATACCAATAGATAATGAT
6310 6320 6330 6340 6350 6360 6370
580 590 600 610 620 630 640
ACTACCACCTATACTGTTGACAAGTTGTAAACACCTCAGTCATTACACAGGCCGTCCAAAGGTATCCTTGAG
|||||
ACTACCACCTATACTGTTGACAAGTTGTAAACACCTCAGTCATTACACAGGCCGTCCAAAGGTATCCTTGAG
6380 6390 6400 6410 6420 6430 6440 6450
650 660 670 680 690 700 710 720
CCAATTCCGATACATTATTCGCCCGGCTGGTTTGCATTCTAAATGTAATAATAAGACGTTCAATGGA
|||||
CCAATTCCGATACATTATTCGCCCGGCTGGTTTGCATTCTAAATGTAATAATAAGACGTTCAATGGA
6460 6470 6480 6490 6500 6510 6520
730 740 750 760 770 780 790
ACAGGACCATGTACAATGTCAACACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTG
|||||
ACAGGACCATGTACAATGTCAACACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTG
6530 6540 6550 6560 6570 6580 6590
800 810 820 830 840 850 860
CTGTTGAATGCGAGTCTAGCAAAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACC
|||||
CTGTTGAATGCGAGTCTAGCAAAAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACC
6600 6610 6620 6630 6640 6650 6660
870 880 890 900 910 920 930
ATAATACTACACCTGAAACCAATCTGTAGAAATTAAATTGTACAAAGACCCAAACAACAATACAAGAAAAAGTATC
|||||
ATAATACTACACCTGAAACCAATCTGTAGAAATTAAATTGTACAAAGACCCAAACAACAATACAAGAAAAAGTATC
6670 6680 6690 6700 6710 6720 6730
940 950 960 970 980 990 1000
CGTATCCAGAAGGGGACGAGAGAGCAATTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGT
|||||
CGTATCCAGAAGGGGACGAGAGAGCAATTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGT
6740 6750 6760 6770 6780 6790 6800 6810

1010 1020 1030 1040 1050 1060 1070 1080
 AACATTAAGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAATTAAAGAGAACAAATTGGAAAT
 : : : : : : : :
 AACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAATTAAAGAGAACAAATTGGAAAT
 6020 6030 6040 6050 6060 6070 6080
 1090 1100 1110 1120 1130 1140 1150
 AATAAACAAATAATCTTAAAGCAATCCTCAGGAAGGGACCCAGAAATTGTAACGCACAGTTTAATTGTGGA
 : : : : : : : :
 AATAAACAAATAATCTTAAAGCAATCCTCAGGAAGGGACCCAGAAATTGTAACGCACAGTTTAATTGTGGA
 6090 6100 6110 6120 6130 6140 6150
 1160 1170 1180 1190 1200 1210 1220
 GGGCAATTTCGACTGTAATTCAACACAACCTGTTAATAGTACTTGGTTAATAGTACTTGGAGTACTGAA
 : : : : : : : :
 GGGCAATTTCGACTGTAATTCAACACAACCTGTTAATAGTACTTGGTTAATAGTACTTGGAGTACTGAA
 6160 6170 6180 6190 6200 6210 6220
 1230 1240 1250 1260 1270 1280 1290
 GGTCAAATACACTGAAAGCAATGCAACACTCCCAGCAGAATAAAACAATTATAAACATGTGG
 : : : : : : : :
 GGTCAAATACACTGAAAGCAATGCAACACTCCCAGCAGAATAAAACAATTATAAACATGTGG
 7030 7040 7050 7060 7070 7080 7090
 1300 1310 1320 1330 1340 1350 1360
 CAGAAAGTAGGAAAGCAATGCAACACTCCCAGCAGAATAAAACAATTATAAACATGTGG
 : : : : : : : :
 CAGAAAGTAGGAAAGCAATGCAACACTCCCAGCAGAATAAAACAATTATAAACATGTGG
 7100 7110 7120 7130 7140 7150 7160 7170
 1370 1380 1390 1400 1410 1420 1430 1440
 CTGCTATTAAAGAGATGGTAAATAACAAACATEGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATG
 : : : : : : : :
 CTGCTATTAAAGAGATGGTGGTAAATAACAAACATEGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATG
 7180 7190 7200 7210 7220 7230 7240
 1450 1460 1470 1480 1490 1500 1510
 AGGGCAAAATTAGAAAGTGAATTATAAAATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACC
 : : : : : : : :
 AGGGCAAAATTAGAAAGTGAATTATAAAATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACC
 7250 7260 7270 7280 7290 7300 7310
 1520 1530 1540 1550 1560 1570 1580
 AAGGCAAAAGAGAGAGTGCGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCTGGGTTCTG
 : : : : : : : :
 AAGGCAAAAGAGAGAGTGCGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTTCCTGGGTTCTG
 7320 7330 7340 7350 7360 7370 7380
 1590 1600 1610 1620 1630 1640 1650
 GGAGCAAGCAGGAGAGCACTATGGGCGCAACGGTCAATGACGGCTGACGGTACAGGCCAGACAATTATTGTCTGGT
 : : : : : : : :
 GGAGCAAGCAGGAGAGCACTATGGGCGCAACGGTCAATGACGGCTGACGGTACAGGCCAGACAATTATTGTCTGGT
 7390 7400 7410 7420 7430 7440 7450
 1660 1670 1680 1690 1700 1710 1720
 ATATGCAAGCAAGCAACAAATTGCTGAGGGCTATTGAGGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG
 : : : : : : : :
 ATATGCAAGCAAGCAACAAATTGCTGAGGGCTATTGAGGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG
 7460 7470 7480 7490 7500 7510 7520 7530
 1730 1740 1750 1760 1770 1780 1790 1800
 GGCATCAAGCAAGCTCCAGGCAAGAAATCCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGATT
 : : : : : : : :
 GGCATCAAGCAAGCTCCAGGCAAGAAATCCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGATT
 7540 7550 7560 7570 7580 7590 7600

1810	1820	1830	1840	1850	1860	1870
TGGGATTCTCTGAAAGACTCATTCGACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAATCTCTG						
TGGEGETTGTCTGGAAAAGCTCATTCGACCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAATCTCTG						
7610	7620	7630	7640	7650	7660	7670
1880	1890	1900	1910	1920	1930	1940
GAACAGATTTGAAATANCATGACCTGGATGGAGTGGGACAGAGAAATTACAATTACACAAGCTTAATACAT						
GAACAGATTTGAAATANCATGACCTGGATGGAGTGGGACAGAGAAATTACAATTACACAAGCTTAATACAT						
7680	7690	7700	7710	7720	7730	7740
1950	1960	1970	1980	1990	2000	2010
TCCTTAATTGAAGAATLGCARAAACCGAACGAAAAAGAATGAACAAGAATTATTGGAATTAGATAAAATGGGCA						
TCCTTAATTGAAGAATLGCARAAACCGAACGAAAAAGAATGAACAAGAATTATTGGAATTAGATAAAATGGGCA						
7750	7760	7770	7780	7790	7800	7810
2020	2030	2040	2050	2060	2070	2080
AGTTTGTTGAGAATTTGTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATAATGATAGTAGGAGGC						
AGTTTGTTGAGAATTTGTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATAATGATAGTAGGAGGC						
7820	7830	7840	7850	7860	7870	7880
2090	2100	2110	2120	2130	2140	2150
TTGGTAAAGAATAGTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAACCATTAA						
TTGGTAAAGAATAGTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAACCATTAA						
7900	7910	7920	7930	7940	7950	7960
2170	2180	2190	2200	2210	2220	2230
TCGTTTCAAGCACCCTCCCAACCCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAG						
TCGTTTCAAGCACCCTCCCAACCCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAG						
7970	7980	7990	8000	8010	8020	8030
2240	2250	2260	2270	2280	2290	2300
AGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTAGCACTTATCTGGGACGATCTGCGGAGCCTT						
AGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTAGCACTTATCTGGGACGATCTGCGGAGCCTT						
8040	8050	8060	8070	8080	8090	8100
2310	2320	2330	2340	2350	2360	2370
GTGGCTCTTCAAGCTAACCGCCTTGAGAGACTTACTCTTGGATTGTAACGAGGATTGTGGAACCTCTGGGACG						
GTGGCTCTTCAAGCTAACCGCCTTGAGAGACTTACTCTTGGATTGTAACGAGGATTGTGGAACCTCTGGGACG						
8110	8120	8130	8140	8150	8160	8170
2380	2390	2400	2410	2420	2430	X
CAGGGGGTGGGAAAGCCCTCAAAATAATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACCTAAAG						
CAGGGGGTGGGAAAGCCCTCAAAATAATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACCTAAAG						
8180	8190	8200	8210	8220	8230	8240

2. KUNZ-158-CL334 SEQ

HIVMN05 Human immunodeficiency virus type 1, isolate MN, c

LOCUS	HIVMN05	5738 bp ss-RNA	VRL	15-JUN-1989
DEFINITION		Human immunodeficiency virus type 1, isolate MN, complete genome.		
ACCESSION	MJ7449			
KEYWORDS				
SOURCE		Human immunodeficiency virus type 1 (HIV-1), isolate MN, proviral DNA.		
ORGANISM		Human immunodeficiency virus type 1 Mammalian ss-RNA enveloped viruses: Retroviridae		

REFERENCE Lentivirinae.
 AUTHORS i (bases 1 to 9738)
 Guringo, C., Guo, H.-G., Franchini, G., Aldovini, A., Collalti, E.,
 Farrell, K., Wong-Staal, F., Gallo, R. C. and Reitz, M. S. Jr.
 TITLE Envelope sequences of two new United States HIV-1 isolates
 JOURNAL Virology 164, 531-536 (1988)
 STANDARD full staff_review
 COMMENT Sequence kindly provided in computer readable form by Marv Reitz,
 N.C.I., Bethesda, MD, 20892 U.S.A.
 The MN isolate was taken from a pediatric AIDS patient in 1984.
 The pol coding sequence shows an in-frame stop codon at position
 3783; the nef protein is prematurely truncated at position 9357.
 The vpu protein is prematurely truncated at position 6142.

FEATURES	from	to/span	description		
pept	787	2307	gag polyprotein		
pept. ps <	2031	5111	pol polyprotein (NH2-terminus uncertain; AA at 2031; in-frame stop codon at 3783)		
pept	5056	5634	vif protein		
pept	5574	5864	vpr protein		
pept	5845	6058	tat protein, exon 2 (first expressed exon)		
	8396	8486	tat protein, exon 3 (AA at 8397)		
pept	5984	6058	rev protein, exon 2 (first expressed exon)		
	8396	8667	rev protein, exon 3 (AA at 8398)		
pept	6073	6144	vpu protein (premature termination)		
pept	6238	8908	envelope polyprotein		
pept	6811	9358	nef protein (premature termination at 9357 relative to other HIV-1 sequences)		
pre-msg	454	9655	genomic mRNA		
pre-msg	454	9655	tat, rev, nef subgenomic mRNA		
IVS	740	5791	tat, rev, nef subgenomic mRNA intron 1		
IVS	6060	8385	tat cds intron 2		
IVS	6060	8395	rev cds intron 2		
IVS	6060	8395	tat, rev, nef subgenomic mRNA intron 2		
LTR	1	633	5' LTR		
LTR	9106	9738	3' LTR		
rpt	453	550	R repeat 5' copy		
rpt	3658	9655	R repeat 3' copy		
binding	376	385	Spl binding site III		
binding	387	396	Spl binding site II		
binding	398	407	Spl binding site I		
binding	635	652	primer (Lys-tRNA) binding site		
site	3783	3785	pol cds in-frame stop codon		
signal	9631	9636	mRNA polyadenylation signal		
BASE COUNT	3463	a 1769	c 2344	g 2162	t

ORIGIN Left end of viral genome

Initial Score = 596 Optimized Score = 2203 Significance = 0.00
 Residue Identity = 90% Matches = 2233 Mismatches = 170
 Gaps = 58 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
ATGAGAGTCAAGGAGAAATATCA	CACTTG	TGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA					
ATGAGAGTCAAGGAGAAATTAT-CAG-CACTGGTGGGGATGGGGCACGATGCTCCTTGGGTTA							
6240	6250	6260	6270	6280	6290	6300	

80	90	100	110	120	130	140
TTGATGATCTGTA	GTCACAGAAAAATTGTGGGT	CACAGTCTATTATGGGGTACCTGTGTGGAAAGGAAGCA				
TTAATGATCTGTA	GTCACAGAAAAATTGTGGGT	CACAGTCTATTATGGGGTACCTGTGTGGAAAGGAAGCA				
6310	6320	6330	6340	6350	6360	6370

150	160	170	180	190	200	210
ACCACCACTCTATT	TTGCA	TGAGATGCTAAAGCATATGATA	CAGAGGGTACATAATGTTGGGCCACACAT			
ACCACCACTCTATT	TTGCA	TGAGATGCTAAAGCATATGATA	CAGAGGGTACATAATGTTGGGCCACACAA			

6380 6390 6400 6410 6420 6430 6440 6450
220 230 240 250 260 270 280
GCCCTGTGTAACCCACACADACCCCCAACCCACAGAAAGTAGTATTGGTAAATGTGACAGAAAAATTAAACATGTGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
GCCCTGTGTAACCCACACADACCCCCAACCCACAGAAAGTAGTATTGGTAAATGTGACAGAAAAATTAAACATGTGG
6460 6470 6480 6490 6500 6510 6520

290 300 310 320 330 340 350 360
AAAAATGACATGGTAGAACAGATGCATGAGGAATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
AAAATAAACATGGTAGAACAGATGCATGAGGAATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
6530 6540 6550 6560 6570 6580 6590

370 380 390 400 410 420
AAATTAACCCCACTCTGTGTTAGTTAAAGTCACTGATTGGGAATGCTACTAATACCAATACTAG----
|||||:|||||:|||||:|||||:|||||:|||||:
AAATTAACCCCACTCTGTGTTACTTTAAATTCACTGATTGGGAATACTACTAATACCAATAATAGTACT
6600 6610 6620 6630 6640 6650 6660

430 440 450 460 470 480 490
--TAATAACAAATAAGTAGTAAAGCGGGGAAATGATGATGGAGAAAGGAGAGATAAAAAACTGCTCTTCATAATAC
|||||:|||||:|||||:|||||:|||||:
GCTAATAACAAATAAGTAGTAAAGCGGGGAAACAATAAGG-GA--GGAGAAATGAAAAACTGCTCTTCATAATAC
6670 6680 6690 6700 6710 6720 6730

500 510 520 530 540 550 560 570
AGCACAAAGATAAGAGATAAGGTGCAAGAAAGAAATATGCATTTTTATAAACTTGATATAATACCAATAGAT
|||||:|||||:|||||:|||||:
ACCACAAACATAAGAGATAAGATEGCAAGAAAGAAATATGCACCTCTTATAAACTTGATATAGTATAAGAT
6740 6750 6760 6770 6780 6790 6800

580 590 600 610 620 630 640
AATGATACTACCAACTATACTTACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCTAAAGGTATCC
|||||:|||||:|||||:
AATGATACTACCAACTATACTTACAAGTTGTAACACCTCAGTCATTACACAGCTTGTCCAAAGATATCC
6810 6820 6830 6840 6850 6860 6870 6880

650 660 670 680 690 700 710
TTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTGCATTCTAAATGTAATAAGACGTTTC
|||||:|||||:
TTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTGCATTCTAAATGTAACGATAAAAGTTTC
6690 6800 6910 6920 6930 6940 6950

720 730 740 750 760 770 780
AATGGAACAGGACCATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACT
|||||:
AGTCGAAAGAGATCATGTAATAATCTCACGACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACT
6860 6870 6880 6890 7000 7010 7020

790 800 810 820 830 840 850
CAACTGCTGTTAAATGCGAGTCTACCAAGAGAGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCT
|||||:
CAACTGCTGTTAAATGCGAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGAGAATTTCACGATAATGCT
7030 7040 7050 7060 7070 7080 7090

860 870 880 890 900 910 920 930
AAAAACCTATAAGTACAGCTGAAACCAATCTGTAGAAATTAAATTGTACAAGACCCAACAACAATACAAGAAAA
|||||:
AAAACCATCATAGTACATCTGAAATCTGTACAATTAAATTGTACAAGACCCAACACTACAATAAAAGAAAA
7100 7110 7120 7130 7140 7150 7160

940 950 960 970 980 990
AGTATCCGTTACGAGGGGACCCAGGGAGASCATTGTACAATAGGAA---AATAGGAAATATGAGACAA
|||:
AGGAGTACAATGTTGAGGAGAGCATTTATACAAACAAAAAATATAATGGAACTATAAGACAA

7170	7180	7190	7200	7210	7220	7230	
1000	1010	1020	1030	1040	1050	1060	1070
GCACATTGTAACATTAGTAGGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAATTAAAGAGAACAA							
GCACATTGTAACATTAGTAGGCAAAATGCAATGCCACTTTAAGACAGATAGCTAGCAAATTAAAGAGAACAA							
7240	7250	7260	7270	7280	7290	7300	
1080	1090	1100	1110	1120	1130	1140	
TTTGGAAATAATAAAACAATACTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGACAGTTT							
TTT---AAGAATAAAACAATACTGTCTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTT							
7310	7320	7330	7340	7350	7360	7370	
1150	1160	1170	1180	1190	1200	1210	
AATTGTTGGAGGGGAATTTTCTACTGTAATTCAACACAACGTGTTAATAGTACTTGG---TTTAATAGTACT							
AATTGTTGGAGGGGAATTTTCTACTGTAATACTACACCACGTGTTAATAGTACTTGGAAATGGTAATAACT							
7380	7390	7400	7410	7420	7430	7440	
1220	1230	1240	1250	1260	1270	1280	
TGG---AGTACTGAAAGGTCAAAATAACACTGAAAGGAAGTGACACACAATCACACTCCCATGCAGAATAAAACAA							
TGGAATAATACTACAGGGGTCAAAATAACAAT-----ATCACACTTCATGCAAAATAAAACAA							
7450	7460	7470		7480	7490	7500	
1290	1300	1310	1320	1330	1340	1350	
TTTATAAACATGTGGCAGGAAGTAGGCAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCA							
ATTATAAACATGTGGCAGGAAGTAGGAAAGCAATGTATGCCCTCCCATGAGGACAAATTAGATGTTCA							
7510	7520	7530	7540	7550	7560	7570	
1360	1370	1380	1390	1400	1410		
TCAAATATTACAGGGCTACTATTACAACAGATGGTGGTAAGGACACGGACACGAACGACACCGAGATCTTC							
TCAAATATTACAGGGCTACTATTACAACAGATGGTGGTAAGGACACGGACACGAACGACACCGAGATCTTC							
7580	7590	7600	7610	7620	7630	7640	
1420	1430	1440	1450	1460	1470	1480	1490
AGACCTGGAGGAAGGAGATAATGAGGGAGACAATTGGAGAGTGAATTATAAAATAAAGTAGTAAAAATTGAA							
AGACCTGGAGGAAGGAGATAATGAGGGAGACAATTGGAGAGTGAATTATAAAATAAAGTAGTAAACAATTGAA							
7650	7650	7670	7680	7690	7700	7710	7720
1500	1510	1520	1530	1540	1550	1560	
CCATTAGGAGTAGCACCCACCAAGGCAAAAGAGAAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGA							
CCATTAGGAGTAGCACCCACCAAGGCAAAAGAGAAAGAGTGGTGCAGAGAGAAAAAGAGCAGCG---ATAGGA							
7730	7740	7750	7760	7770	7780		
1570	1580	1590	1600	1610	1620	1630	
GCTTGTTCCTTGGGTTCTGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAG							
GCTCTGTTCCTTGGGTTCTTGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAG							
7790	7800	7810	7820	7830	7840	7850	7860
1640	1650	1660	1670	1680	1690	1700	
GCCAGACAATTATTGTCTGGTATAGTGCAAGCAGCAGAACAAATTGCTGAGGGCTATTGAGGCGCAACAGCAT							
GCCAGACAATTATTGTCTGGTATAGTGCAACAGCAGAACAAATTGCTGAGGGCCATTGAGGCGCAACAGCAT							
7870	7880	7890	7900	7910	7920	7930	
1710	1720	1730	1740	1750	1760	1770	
CTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAACATCCTGGCTGTGGAAAGATACCTAAAG							
ATGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCTGGCTGTGGAAAGATACCTAAAG							

7840	7850	7860	7870	7880	7890	8000	
1780	1790	1800	1810	1820	1830	1840	1850
GATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAACTCATTTGCACCACTGCTGCCTTGAATGCT							
GATCAACAGCTCCTGGGGTTGCTCTGGAAACTCATTTGCACCACTACTGTGCCTTGAATGCT							
8010	8020	8030	8040	8050	8060	8070	
1860	1870	1880	1890	1900	1910	1920	
AGTTGGAGTAATAAACTCTGGAACAGATTGGAAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAAAC							
AGTTGGAGTAATAAAATCTCTGGATGATATTGGAAATAACATGACCTGGATGCAAGTGGGAAAGAGAAATTGAC							
8080	8090	8100	8110	8120	8130	8140	
1930	1940	1950	1960	1970	1980	1990	
AATTACACAAGCTTAATACATTCTTAATTGAAGAACGAAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTA							
AATTACACAGCTTAATATACTCATTACTAGAAAAATCGCAAACCCACAAGAAAAGAATGAACAAGAATTA							
8150	8160	8170	8180	8190	8200	8210	8220
2000	2010	2020	2030	2040	2050	2060	
TTGGAATTAGATAAAATGGGCAAGTTGTGGAATTGGTTAACATAACAAATTGGCTGTGGTATATAAAAATA							
TTGGAATTGGATAAAATGGGCAAGTTGTGGAATTGGTTGACATAACAAATTGGCTGTGGTATATAAAAATA							
8230	8240	8250	8260	8270	8280	8290	
2070	2080	2090	2100	2110	2120	2130	
TTCATAATGATAGTAGGAGGCCTTGGTAGGTTAACAGATAGTTTGCTGTACTTCTATAGTGAATAGAGTT							
TTCATAATGATAGTAGGAGGCCTTGGTAGGTTAACAGATAGTTTGCTGTACTTCTATAGTGAATAGAGTT							
8300	8310	8320	8330	8340	8350	8360	
2140	2150	2160	2170	2180	2190	2200	2210
AGGCAGGGATATTCAACCATTATCGTTTCAAGACCCACCTCCCAACCCCCGAGGGGACCCGACAGGCCGAAGGA							
AGGCAGGGATACTCACCATTCGTTGCAGACCCGCCCCCAGTTCCGAGGGGACCCGACAGGCCGAAGGA							
8370	8380	8390	8400	8410	8420	8430	
2220	2230	2240	2250	2260	2270	2280	
ATAGAAGAAAGAARGTGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATC							
ATCGAAGAAAGAAGGTGAGAGAGAGACAGAGACACATCCGGTCGATTAGTGCATGGATTCTTAGCAATTATC							
8440	8450	8460	8470	8480	8490	8500	
2290	2300	2310	2320	2330	2340	2350	
TGGGACGATCTGGGAGCCTTGTCCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAG							
TGGGTCGACCTGGGAGGCC-TGGTCCCTCTTCAGCTACCACAC---AGAGACTTACTCTTGATTGAGCGAG							
8510	8520	8530	8540	8550	8560	8570	
2360	2370	2380	2390	2400	2410	2420	
GATTGTGGAACCTCTGGGAGCGCAGGGGTGGGAAGGCCCTCAAATATTGGTGGAATCTCCTACAGTATTGGAG							
GATTGTGGAACCTCTGGGAGCGCAGGGGTGGGAAGTCCTCAAATATTGGTGGAATCTCCTACAGTATTGGAG							
8580	8590	8600	8610	8620	8630	8640	
2430	2440						
TCAGGAACTAAAG							
TCAGGAACTAAAG							
8650	8660	X					

700 710 720 730 740 750 760
290 300 310 320 330 340 350 360
AAAAATGACATGGTAGAACAGATECATGAGGAATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AAAAATAACATGGTAGAACAGATECATGAGGAATAATCAGCTTATGGGATCAAAGCCTAAAGCCATGTGTA
770 780 790 800 810 820 830

370 380 390 400 410 420 430
AAATTAAACCCCACCTCTGTGTTAGTTAAAGTGCACTGATTTGGGAAATGCTACTAATACCAACTA--GTA
||| ||||||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAACTAACCCCCACCTCTGTGTTACTTTAAATTGCACTGATTTGAATACTAATAACTACTAATAACTGAA
840 850 860 870 880 890 900 910

440 450 460 470 480 490 500
ATACCAATAAGTAGTTAGCGGGGAAATGATGATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCATATCAGCA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTATCAATAATAGTAGTTGGGAAACAACG--GGGTAAAGGAGAAATGAGAAACTGTTCTTCATATCACCA
920 930 940 950 960 970 980

510 520 530 540 550 560
CAAGNATAAGGGTAAGGTCAGAAAAGAATATGCATTTTATAAAACTTGATATAATACCAATAG-----
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CAAGCATAAGGATAAGGTGCAAGAGAGAATATGCATTGTTATAAAACTTGATGTAGAACCAATAGATGATA
990 1000 1010 1020 1030 1040 1050

570 580 590 600 610 620 630
ATAATGATACTACCCAGC-----TATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCCTGTC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATAAAAATACTACCAACAAACACCAAATAAGGTTGATAAAATTGTAACACCTCAGTCATTACACAGGCCCTGTC
1050 1070 1080 1090 1100 1110 1120

640 650 660 670 680 690 700
CAAAGGTATCCTTGAAGCCAATTCCATACATTATTGTGCCCCGGCTGGTTTGCATTCTAAAATGTAATA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CAAAGGTATCCTTGAAGCCAATTCCATACATTATTGTACCCGACTGGTTTGCACCTCTAAAGTGTAAACG
1130 1140 1150 1160 1170 1180 1190

710 720 730 740 750 760 770
ATAAGACCTTCAATGGAACAGGACCATEACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATAAGAACCTTCAATGGAACAGGACCATEACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAG
1200 1210 1220 1230 1240 1250 1260

780 790 800 810 820 830 840
TAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGCAATTTC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGAAAATTTC
1270 1280 1290 1300 1310 1320 1330 1340

850 860 870 880 890 900 910
CAGACAATGCTAAACCCATAATAGTACAGCTGAACCAATCTGTAGAAATTAAATTGTACAAGACCCAACAAACA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CGAACAAATGCTAAACCCATAATAGTACAGCTGAATGTATCTGTAGAAATTAAATTGTACAAGACCCAACAAACC
1350 1360 1370 1380 1390 1400 1410

920 930 940 950 960 970 980
ATACAAGAAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTGTTACAATAGGAAAAA---TAGGAA
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATACAAGAAAAG---GGTAAAC---GCTAGGGACCAAGGGAGAGTATGGTATACAACAGGGAGAAATACTAGGAA
1420 1430 1440 1450 1460 1470

990 1000 1010 1020 1030 1040 1050 1060
ATATGAGACAACGACATTGTAAACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAAT
||| ||| ||| ||| ||| ||| ||| ||| |||
ATATGAGACAACGACATTGTAAACATTAGTAGAGCAAAATGCAATGCCACTTTACAACAGATAGCTAGCAAACCT

	2270	2280	2290	2300	2310	2320	2330
1840	1850	1860	1870	1880	1890	1900	
TGTGCTTGGAACTCTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGAATAACATGACCTGGATGGAGTG							
TGTGCTTGGAACTCTAGTTGGAGTAATAAAACTCTGGATCAGATTGGAATAACATGACCTGGATGGAGTG							
18340	18350	18360	18370	18380	18390	18400	
1910	1920	1930	1940	1950	1960	1970	
GGACAGAGAAATTAAACAATTACACAAAGCTTAATACATTCCCTAATTGAAGAATCGCAAAACCCAGCAAGAAAA							
GGACAGAGAAATTAAACAATTACACACACTTAATACACTTTAATTGAAGAATCGCAAAACCCAAACAAGAAAA							
2410	2420	2430	2440	2450	2460	2470	
1980	1990	2000	2010	2020	2030	2040	2050
GAATGAACAAAGAATTATTGGAATTAGATAAAATGGGCAAGTTGTGGATTGGTTAACATAACAAATTGGCT							
GAATCAACAGGAACTATTGCAATTAGATAAGTGGGCAAGTTGTGGACTTGGTCTGACATAACAAATGGCT							
2480	2490	2500	2510	2520	2530	2540	
2060	2070	2080	2090	2100	2110	2120	
GTGGATATAAAAAATATTCTATAATGATAGTAGGAGGGCTTGGTAGGTTAACAGATAGTTTGCTGTACTTC							
GTGGATATAAAAAATATTCTATAATGATAGTAGGAGGGCTTGGTAGGTTAACAGATAGTTTGCTGTGCTTC							
2550	2560	2570	2580	2590	2600	2610	2620
2130	2140	2150	2160	2170	2180	2190	
TATAATGAAATAGCTTGGCAGGAACTTCACCATTATCGTTTCAGACCCACCTCCAAACCCCGAGGGGGACC							
TATAATGAAATAGCTTGGCAGGAACTTCACCATTATCGTTTCAGACCCACCTCCAAACCCCGAGGGGGACC							
2630	2640	2650	2660	2670	2680	2690	
2200	2210	2220	2230	2240	2250	2260	
CGACAGGCCGAAAGCAATAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGG							
CGACAGGCCGAAAGGAACCEAABAGGGAGGTGGAGAGAGAGGGCAGAGACGGATCCACTCGATTAGTGCATGG							
2700	2710	2720	2730	2740	2750	2760	
2270	2280	2290	2300	2310	2320	2330	
ATCCTTAACCACTTATCTGGGACGGATCTCGGGAGCCTTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAC							
CTTCTTAACCACTTGTGTGGGACGGATCTCGGGAGCC-TGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAC							
2770	2780	2790	2800	2810	2820	2830	
2340	2350	2360	2370	2380	2390	2400	2410
TCTTGTATTGTAACGGAGGATTCTGAACTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAAATC							
TCTTGTATTGTAACGGAGGATTCTGAACTTCTGGGACGCAGGGGGTGGGAAGTCCCAAATATTGGTGGAAATC							
2840	2850	2860	2870	2880	2890	2900	
2420	2430	2440					
TCCTACAGTATTGGAGTCAGGAACCTAAAG							
TCCTACAGTATTGGAGTCAGGAACCTAAAG							
2910	2920	2930	X				

4. KUNZ-158-CL33. SEE

HIVPV22 Human immunodeficiency virus type 1, isolate PV22,

LOCUS HIVPV22 8770 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate PV22, complete genome (HIV/HTLV-III proviral DNA).
 ACCESSION K02083
 KEYWORDS TAP protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein;

	provirus; rsv gene; reverse transcriptase; tat gene; trans-activator.		
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate PV22 (from H9-derived family), proviral DNA.		
ORGANISM	Human immunodeficiency virus type 1 Viiidae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.		
REFERENCE	1 (bases 1 to 8770; revised sequence, personal communication)		
AUTHORS	Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A. and Capon, D. J.		
TITLE	Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus		
JOURNAL	Nature 313, 450-458 (1985)		
STANDARD	full staff_review		
REFERENCE	2 (bases 2111 to 2112; revises [1])		
AUTHORS	Muesing, M. A.		
JOURNAL	Unpublished (1987) Whitehead Inst Cambridge, Mass		
STANDARD	full staff_review		
COMMENT	This sequence for a H9/HTLV-III virus was determined from one complete proviral clone [1]. Additionally, several cDNA clones of the viral RNA were sequenced for comparison with the entire proviral sequence. The differences between cDNA and proviral DNA are extensive and are listed in the Sites Table as variations. The authors believe that the variations may be due in part to different strains in the H9/HTLV-III cell line, because it was established by infection with material from several AIDS patients. With the addition of g at 2111, gag cds and pol cds are very close to those of HM22, ERU, and related HIV viruses. For details and other references pertaining to Sites and Features, see the HIV reference entry.		
FEATURES	from to/span description		
pept	789	2337	gag polyprotein precursor
pept	< 2084	5141	pol polyprotein (NH2-terminus uncertain; AA at 2084)
pept	5086	5684	vif protein
pept	5604	5840	vpr protein
pept	5876	6080	tat protein, exon 2 (first expressed exon)
pept	8421	8466	tat protein, exon 3 (AA at 8422)
pept	8015	8090	rev protein, exon 2 (first expressed exon)
pept	8421	8686	rev protein, exon 3 (AA at 8423)
pept	6107	6352	vpu protein
pept	6267	8837	envelope polyprotein
pept	6939	9459	nef protein
pre-msg	464	5670	genomic mRNA
pre-msg	464	5678	tat, rev, nef subgenomic mRNA
IVS	753	5822	tat, rev, nef subgenomic mRNA intron 1
IVS	6091	8420	tat cds intron 2
IVS	6091	8420	rev cds intron 2
IVS	6091	8420	tat, rev, nef subgenomic mRNA intron 2
LTR	1C	643	5' LTR
LTR	9128	9761	3' LTR
rpt	463	560	R repeat 5' copy
rpt	9581	9676	R repeat 3' copy
binding	386	385	Sp1 binding site III
binding	397	406	Sp1 binding site II
binding	406	417	Sp1 binding site I
binding	645	662	primer (Lys-tRNA) binding site
variant	510	510	a in provirus; g in cDNA [1]
variant	575	575	g in provirus; a in cDNA [1]
revision	2111	2112	gg in [2]; g in [1]
variant	5716	5716	g in provirus; a in cDNA [1]
variant	5892	5882	a in provirus; g in cDNA [1]
variant	6007	6007	c in provirus; t in cDNA [1]
variant	6047	6047	c in provirus; g in cDNA [1]
variant	6051	6051	c in provirus; a in cDNA [1]
variant	6055	6057	agg in provirus; gaa in cDNA [1]

variant	6108	6108	t in provirus; c in CDNA [1]
variant	6120	6120	a in provirus; c in CDNA [1]
variant	6126	6126	gc in provirus; gtaac in CDNA [1]
variant	6136	6136	a in provirus; c in CDNA [1]
variant	6236	6236	t in provirus; a in CDNA [1]
variant	6352	6352	g in provirus; a in CDNA [1]
variant	6760	6760	t in provirus; a in CDNA [1]
variant	7090	7050	c in provirus; t in CDNA [1]
variant	7100	7100	a in provirus; g in CDNA [1]
variant	7134	7136	ca in provirus; ac in CDNA [1]
variant	7183	7184	gt in provirus; aa in CDNA [1]
variant	7199	7199	a in provirus; g in CDNA [1]
variant	7204	7205	aa in provirus; gc in CDNA [1]
variant	7303	7303	a in provirus; c in CDNA [1]
variant	7511	7511	a in provirus [1]; c in CDNA [1]
variant	7533	7533	t in provirus [1]; a in CDNA [1]
variant	7586	7586	c in provirus [1]; t in CDNA [1]
variant	7648	7648	a in provirus [1]; g in CDNA [1]
variant	8138	8138	a in provirus; c in CDNA [1]
variant	8143	8143	t in provirus; c in CDNA [1]
variant	8222	8222	g in provirus; a in CDNA [1]
variant	8269	8269	a in provirus [1]; g in CDNA [1]
variant	8285	8295	g in provirus [1]; t in CDNA [1]
variant	8376	8376	a in provirus [1]; g in CDNA [1]
variant	8381	8381	a in provirus [1]; g in CDNA [1]
variant	8476	8476	a in provirus [1]; g in CDNA [1]
variant	8869	8869	a in provirus [1]; g in CDNA [1]
variant	8879	8879	c in provirus; t in CDNA [1]
variant	8880	8880	a in provirus; c in CDNA [1]
variant	8899	8899	c in provirus [1]; a in CDNA [1]
variant	9031	9031	a in provirus [1]; g in CDNA [1]
variant	9291	9291	t in provirus [1]; g in CDNA [1]
variant	9295	9295	g in provirus [1]; t in CDNA [1]
variant	9303	9303	g in provirus [1]; a in CDNA [1]
variant	9548	9548	g in provirus [1]; c in CDNA [1]
signal	9654	9659	mRNA polyadenylation signal
prov	10	9781	HIV-1 proviral DNA
cell	1	5	human cellular DNA
cell	9762	9770	human cellular DNA
BASE COUNT	3438	a 1786 c 2376 g 2172 t	
ORIGIN	482	bp upstream of BglII site.	

Initial Score = 1877 Optimized Score = 2180 Significance = 0.00
Residue Identity = 85% Matches = 2247 Mismatches = 146
Gaps = 106 Conservative Substitutions = 0

AAGAGATGATGATACAGAGGTACATAATGTTGGGC-CACA--CATGCCTGTGACCCACAGACCCCCAACCC-C
 6450 6460 6470 6480 6490 6500
 250 260 270 280 290 300 310
 AAGAGATGATGATTCGTTAAATGTCACAGAAAAATTAAACATGTGGAAAA--ATGACATGGTAGAACAGATG-C
 6510 6520 6530 6540 6550 6560
 320 330 340 350 360 370 380
 ATGGAGATATATGCACTTTCAT-GGATCAAACCTAAAGCCATGTG-TAAAA--TTAACCCCACACTCTGTGT
 6570 6580 6590 6600 6610 6620
 390 400 410 420 430 440
 TAGTTAAAGCTGATTTGG---CGAATGCTACTAAT---ACCAACTAGTAATACCAATAGTAGTA
 6630 6640 6650 6660 6670 6680 6690
 AAATTAACCCCACTGTGTATTTAAAGTGC-ACTGATTTGAAGAATGATACTAATACCAATAGTAGTA
 450 460 470 480 490 500 510
 GCGGAGAAATGATGATGAAAGAGAGATAAAAAACTGCTTTCAATATCAGCACAGNATAAGAGGTA
 6700 6710 6720 6730 6740 6750 6760
 520 530 540 550 560 570 580
 ACGTGCAAAAGATAATGCACTTTTATAAAACTGATATAATACCAATAGATAATGATACTACCAGCTATA
 6770 6780 6790 6800 6810 6820 6830 6840
 590 600 610 620 630 640 650 660
 CGTTGAGAAGTGTGACCTGACCTACATTACACAGGCCCTGTCACAGGTTACCTTGGACCAATTCCCACAC
 6850 6860 6870 6880 6890 6900 6910
 670 680 690 700 710 720 730
 ATTATTTGGCCCGGGCTGGTTGGATTCTAAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
 6920 6930 6940 6950 6960 6970 6980
 740 750 760 770 780 790 800
 CAAATGTCAGCACAGTACATGATCACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
 6990 7000 7010 7020 7030 7040 7050
 810 820 830 840 850 860 870
 GTCTAGCAAGAGAGGGTAGTAATTAGATCTGCCAATTCACAGACAATGCTAAACCAATAGTACAGC
 7060 7070 7080 7090 7100 7110 7120
 880 890 900 910 920 930 940
 TGAACCAATGCTGATGAAATTGATGACAGACCCAAACAACAAATACAAGAAAAAGTATCCGTATCCAGAGGG
 7130 7140 7150 7160 7170 7180 7190 7200
 950 960 970 980 990 1000 1010 1020

GACCCGCGGGGAGCATT GTTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
 7210 7220 7230 7240 7250 7260 7270
 1030 1040 1050 1060 1070 1080 1090
 CAAATGCTATGCCACTTTAACAGATAGCTAGCAAATTAGAGAACAAATTGGAAATAATAAAAACAATAA
 CAAATGAAATAACAUUTTAAACAGATAGATAGCAAATTAGAGAACAAATTGGAAATAATAAAAACAATAA
 7280 7290 7300 7310 7320 7330 7340
 1100 1110 1120 1130 1140 1150 1160
 TCTTTAACCAATCCTAGGAGGGGACCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
 TCTTTAACCAATCCTAGGAGGGGACCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
 7350 7360 7370 7380 7390 7400 7410
 1170 1180 1190 1200 1210 1220 1230
 ACTGAAATTGAAACACACTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
 ACTGAAATTGAAACACACTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
 7420 7430 7440 7450 7460 7470 7480
 1240 1250 1260 1270 1280 1290 1300
 CTGAAAGTGAAGTCACAAATCACACTCCATGCAGAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAA
 CTGAAAGTGAAGTCACAAATCACACTCCATGCAGAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAA
 7490 7500 7510 7520 7530 7540 7550 7560
 1310 1320 1330 1340 1350 1360 1370 1380
 AAGGAAATGATCCCCAACATCAGGAGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA
 AAGGAAATGATCCCCAACATCAGGAGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA
 7570 7580 7590 7600 7610 7620 7630
 1390 1400 1410 1420 1430 1440 1450
 GAGCTGTTGAGTAAACAAATGGTCGGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAAATTGGA
 GAGCTGTTGAGTAAACAAATGGTCGGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAAATTGGA
 7640 7650 7660 7670 7680 7690 7700
 1460 1470 1480 1490 1500 1510 1520
 GAAGTGAAATTATCAATATAAATTTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
 GAAGTGAAATTATCAATATAAATTTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
 7710 7720 7730 7740 7750 7760 7770
 1530 1540 1550 1560 1570 1580 1590
 GAGCTGTTGAGTAAACAAATGGTCGGAGATCTTCAGACCTGGAGGGTTCTGGGAGCAGCAGGAA
 GAGCTGTTGAGTAAACAAATGGTCGGAGATCTTCAGACCTGGAGGGTTCTGGGAGCAGCAGGAA
 7780 7790 7800 7810 7820 7830 7840
 1600 1610 1620 1630 1640 1650 1660
 GCGCTATGCTGAGCTGAAATGGCTGACCGTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
 GCGCTATGCTGAGCTGAAATGGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
 7850 7860 7870 7880 7890 7900 7910 7920
 1670 1680 1690 1700 1710 1720 1730 1740
 AGAACGAAATGAGGAGGGCTTGTGGCGAACAGGATCTGTTGCAACTCACAGTCTGGGAGCAGCAGC
 AGAACGAAATGAGGAGGGCTTGTGGCGAACAGGATCTGTTGCAACTCACAGTCTGGGAGCAGCAGC
 7930 7940 7950 7960 7970 7980 7990
 1750 1760 1770 1780 1790 1800 1810
 AGAACGAAATGAGGAGGGCTTGTGGCGAACAGGATCTGTTGCAACTCACAGTCTGGGAGCAGCAGC

TCCAGGCGAACTTGTGATGGGAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG
 8000 8010 8020 8030 8040 8050 8060
 1820 1830 1840 1850 1860 1870 1880
 GAAACTGTTTGACGACTGCTGGCTTGGAGTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGA
 1890 1900 1910 1920 1930 1940 1950
 ATACGATGTTGGACGACTGCTGGCTTGGAGTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGA
 1960 1970 1980 1990 2000 2010 2020
 AATGGGAAATGGGAAATGGGAAATGGGAAATGGGAAATGGGAAATGGGAAATGGGAAATGGGAAATT
 2030 2040 2050 2060 2070 2080 2090 2100
 GGTGAACTTGTGATGGCTTGGATATAAAATATTCTATAATGATAAGTAGGAGGCTTGGTAGGTTAA
 2110 2120 2130 2140 2150 2160 2170
 GAATAGTGTGGCTTACTTCTAATGATAAGAGTTAGGCAGGGATATTCACTTATCGTTTCAGACCC
 2180 2190 2200 2210 2220 2230 2240
 ACC1CCCAACCCGAGACGGACCCGACAGGCCGAAGGAATAGAAGAAGAAGGGTGGAGAGAGACAGAGACA
 2250 2260 2270 2280 2290 2300 2310
 GATGGATTGATTAATGAAAGGAGCTTGGGACGATCTGGGAGCCTTGTGCCTCTTCAGC
 2320 2330 2340 2350 2360 2370 2380
 TAGTTCGGAGATGGAGAGCTTGGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAA
 2390 2400 2410 2420 2430 X
 GCGGCGAAATATGGGAAATCTGGCTTGGAGTATTGGAGTCAGGAACCTAAAG
 2440 2450 2460 2470 2480 2490
 GCGGCGAAATATGGGAAATCTGGCTTGGAGTATTGGAGTCAGGAACCTAAAG
 2500 2510 2520 2530 2540 2550 2560

5. KUNZ-LIEB-CLUB, BET.

HIV-BH102 Human immunodeficiency virus type 1, isolate BH10,

LOCUS KIVBH102 6632 bp ss-RNA **VRL** 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate BH10, genome.
ACCESSION M15654 K0201C K02008 K02003
KEYWORDS T-cell infection; acquired immune deficiency syndrome; env gene;

SEGMENT	gag gene; long terminal repeat; pol gene; polyprotein; provirus; rev gene; tRNA ^U ; transcriptase; trans-activator.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1) proviral DNA clone BH10.			
ORGANISM	Human immunodeficiency virus type 1 Mammals; S-RNA; enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	[1] (bases 1 to 8832)			
AUTHORS	Rothblat, L., Hezelgrave, W., Patarca, R., Livak, K. J., Starcich, B., Josephson, S., Unruh, E., Rafalski, J. A., Whitehorn, E. A., Blumenthal, K., Ivanoff, L., Fettaway, S. R. Jr., Pearson, M. L., Leiter, U., Papas, T. S., Ghrayeb, J., Chang, N. T., McLeod, C. and Vong-Stael, F.			
TITLE	Complete nucleotide sequence of the AIDS virus, HTLV-III			
JOURNAL	Nature 313, 277-284 (1985)			
STANDARD	full staff review			
COMMENT	<p>The BH10 sequence differs from BH8 and BH5 by 0.9% in the coding regions and 1.8% in the noncoding regions, and the authors of [1] believe that these are stable variants.</p> <p>The HTLV-III genome encodes at least seven proteins: gag, pol, env, tat, rev, 27K antigen and the sor 23K product. The 3' ORF (positions 8153-8773) is truncated in BH10 (stop codon at positions 8502-8524), but reads through in BH8 and other sequences to yield what is now called the 27K antigen.</p> <p>The mechanism for pol gene translation has not been elucidated: a gag-pol fusion protein is possible; splicing or frameshift have not been ruled out. The viral protease would be determined by the region in question.</p> <p>The Tat protein (trans-activator protein, approximately 14 kb) is an effector of an autostimulatory pathway through interaction with a positive control element, the trans-activating responsive sequence (TAR). Tat seems to be a transcriptional control molecule in HTLV-III, but is both that and a post-transcriptional regulatory molecule in HTLV-III. Deletion mutants in the tat gene are incapable of productive replication and exhibit no cytopathic effects in T4 cell lines.</p> <p>In addition to the ~9.4 kb genomic mRNA, subgenomic mRNAs of 7.4, 5.6, 5.0, 4.3, 2.6 and 1.8 kb have been detected.</p>			
FEATURES	start	end	description	
pol	133	1500	gag polyprotein precursor	
pept	133	1407	pol polyprotein (NH2-terminus uncertain; AA at 1407)	
pol	4783	4877	vif protein	
prote	4917	5154	vpr protein	
pol	5139	5433	tat protein, exon 2 (first expressed exon)	
	7734	7749	tat protein, exon 3 (AA at 7735)	
pept	5820	5903	rev protein, exon 2 (first expressed exon)	
	7734	8003	rev protein, exon 3 (AA at 7736)	
pol	6100	6155	vpu protein	
pept	6100	6150	envelope polyprotein	
pol	10121	10703	nef protein, exon 3 (first expressed exon; premature termination)	
pre-meg	<	>	2502	genomic mRNA
pre-meg	<	>	2502	tat, rev, nef subgenomic mRNA
IVS	82	5173	tat, rev, nef subgenomic mRNA intron 1	
IVS	5403	7738	tat cds intron 2	
IVS	5403	7738	rev cds intron 2	
IVS	5403	7738	tat, rev, nef subgenomic mRNA intron 2	
LTR	6443	6532	3' LTR	
rpt	6851	6852	R repeat 3' copy	
BASE COUNT	3057 a	1612 c	2147 g	1976 t
ORIGIN	About 120 bp downstream from HIVBH101.			

Initial Clustal score = 1770 Optimized Score = 2176 Significance = 0.00
 Residue mismatch = 150 Mismatches = 150

X 30 50 70 90 30 40 50 60
ATGAGAAGTCTG-CAGAAGAA-ATGAGACACTTGTGGAGA-TCGGGGTGCGAAATGGGGCAC-CATGCTCCTTGG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTGGGGTGCGAAATGGGGCAC-CATGCTCCTTGG
X 5560 5570 5580 5590 5600 5610
70 90 100 110 120
GATAATTGATGTT---ATGTTGATGCTACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GGTGGAGATGAGGGCACTTGTGATCTGAGTGTGCTACAGAAAAATTGTGGGTACAG
5620 5630 5640 5650 5660 5670 5680 5690
130 140 150 160 170 180
-HGT---GTTGATGCTACAGA-CGACCA-CTCTATTTGTGCATCAGATGCTAAAGCATATGAT
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TCTATTGCTTACGTTTGTGCTACAGAACGACCACTCTATTG-----TGC---ATCAGATGCT
5710 5720 5730 5740 5750
190 200 210 220 230 240
A---GAGAAGTCTG-CAGAAGAA-ATGAGACACTTGTGGAGA-TCGGGGCACACATGGCTG--TGTACCCACAGA-CCCCAACCCAC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAACGATATGAGGGCACTTGTGCTACAGAACGACCACTCTATTG-----TGC---ATCAGATGCT
5760 5770 5780 5790 5800 5810 5820
250 260 270 280 290 300 310
AAAGAGATGAGGGCACTTGTGCTACAGAAAAATTGTGCGAAA--ATGACATGGTAGAACAGATG-C
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
A-GAGGAGATGAGGGCACTTGTGCTACAGAACGACCACTCTATTG-----TGC---ATCAGATGCT
5830 5840 5850 5860 5870
320 330 340 350 360 370 380
ATGAGAAGTCTG-CAGAAGAA-ATGAGACACTTGTGGAGA-TCGGGGCACACATGGCTG-TAAAAA--TTAACCCCACCTCTGTG
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AT---GGGAGGGCACTTGTGCTACAGAACGACCACTCTATTG-----TCAGTTATGGGATCAAAGCCTAAAGCCA---TGTGT
5880 5890 5900 5910 5920 5930
390 400 410 420 430 440
TAGTTTAACTGAGATGATGTTG---GAAATGCTACTAAT---ACCAATACTAGTAATACCAATAGTAGTA
||| ||| ||| ||| ||| ||| ||| ||| |||
AAAGTTTAACTGAGATGATGTTG---GAAATGCTACTAATACCAATAGTAGTA
5940 5950 5960 5970 5980 5990 6000
450 460 470 480 490 500 510
GGCGAGAGATGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATATCAGCACAAGNATAAGAGGTA
||| ||| ||| ||| ||| ||| |||
GGCGAGAGATGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATATCAGCACAAGCATAAGAGGTA
6010 6020 6030 6040 6050 6060 6070 6080
520 530 540 550 560 570 580
AGTGTGAGGAGATGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATATCAGCACAAGNATAAGAGGTA
||| ||| ||| ||| ||| |||
AGTGTGAGGAGATGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATATCAGCACAAGCATAAGAGGTA
6090 6100 6110 6120 6130 6140 6150
590 600 610 620 630 640 650 660
CGTGGAGAGGTTGAGACACCTGAGTCATTACACAGGCCCTGCGAAAGGTATCCTTGAGGCAATTCCCACATAC
||| ||| ||| ||| ||| |||
CGTGGAGAGGTTGAGACACCTGAGTCATTACACAGGCCCTGCGAAAGGTATCCTTGAGGCAATTCCCACATAC
6160 6170 6180 6190 6200 6210 6220
670 680 690 700 710 720 730
ATTGTTGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATGGGAAACAGGACCATGTA
||| ||| ||| |||
ATTGTTGAGGGCACTTGTGCTACAGAACGACCACTCTGTGCTTCAATGGGAAACAGGACCATGTA
||| ||| ||| |||

6210	6215	6220	6225	6230	6235	6240	6245
740	750	760	770	780	790	800	
CAAACTGTCAGTACAGCTTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA							
CAAACTGTCAGCAGTACAGCTTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA							
6300	6310	6315	6320	6325	6330	6335	6340
810	820	830	840	850	860	870	
GTCAGCAAAAGAAGAAGTAACTTAACTGATCTGCCAATTTCACAGACAATGCTAAAACCATAATAGTACAGC							
GTCAGCAAAAGAAGAAGTAACTTAACTGATCTGCCAATTTCACAGACAATGCTAAAACCATAATAGTACAGC							
6370	6380	6390	6400	6410	6420	6430	6440
880	890	900	910	920	930	940	
TGAACCAAATCTGAGAGAGATTAATGAGAAACAAACAAATACAAGAAAAAGTATCCGTATCCAGAGGG							
TGAACCAAATCTGAGAGAGATTAATGAGAAACAAACAAATACAAGAAAAAGTATCCGTATCCAGAGAG							
8450	8460	8470	8480	8490	8500	8510	
950	960	970	980	990	1000	1010	1020
GACCGGGAGAGAGCTTGTGAGAGAGAAAAATAGGAAATATGAGAGACAAGCACATTGTAACATTAGTAGAG							
GACCGGGAGAGAGCTTGTGAGAGAGAAAAATAGGAAATATGAGAGACAAGCACATTGTAACATTAGTAGAG							
6520	6530	6540	6550	6560	6570	6580	
1030	1040	1050	1060	1070	1080	1090	
CAAAATGCAATTCGGCTTAACTGAGAGAGCTAACGAAATTAAAGAGAACATTGGAAATAATAAAACAATAA							
CAAAATGCAATTCGGCTTAACTGAGAGAGCTAACGAAATTAAAGAGAACATTGGAAATAATAAAACAATAA							
6590	6600	6610	6620	6630	6640	6650	
1100	1110	1120	1130	1140	1150	1160	
TCTTTAAGCACTCTGAGGAGGAGCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGAAATTTTCT							
TCTTTAAGCACTCTGAGGAGCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGAAATTTTCT							
6660	6670	6680	6690	6700	6710	6720	
1170	1180	1190	1200	1210	1220	1230	
ACTATAATTCAACAAVACTGTTAATAGTACTTGGTTAATAGTACTTGAGACTGAAAGGGTCAAATAACA							
ACTATAATTCAACAAVACTGTTAATAGTACTTGGTTAATAGTACTTGAGACTAAAGGGTCAAATAACA							
6730	6740	6750	6760	6770	6780	6790	6800
1240	1250	1260	1270	1280	1290	1300	
CTGAAGGAAGTCACACGATCAGCTCCCCTCCCATGCAAGAACATAAAACAATTATAAACATGTGGCAGGAAGTAGGAA							
CTGAAGGAAGTCACACGATCAGCTCCCCTCCCATGCAAGAACATAAAACAATTATAAACATGTGGCAGGAAGTAGGAA							
6810	6820	6830	6840	6850	6860	6870	
1310	1320	1330	1340	1350	1360	1370	1380
AAGCAATGTAATCCCCCTCCCATGCAAGAACATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA							
AAGCAATGTAATCCCCCTCCCATGCAAGAACATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA							
6880	6890	6900	6910	6920	6930	6940	
1390	1400	1410	1420	1430	1440	1450	
GAGATGGTGGTAATACACGATGTTGGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAATTGGA							
GAGATGGTGGTAATACACGATGTTGGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAATTGGA							
6950	6960	6970	6980	6990	7000	7010	
1460	1470	1480	1490	1500	1510	1520	
GAAGTGAATTATATAATATAATGAGTAGTAAATTGAAACCATTAAGGAGTAGCACCCACCAAGGGCAAAGAGAA							
GAAGTGAATTATATAATATAATGAGTAGTAAATTGAAACCATTAAGGAGTAGCACCCACCAAGGGCAAAGAGAA							

7020	7030	7040	7050	7060	7070	7080
1530	1540	1550	1560	1570	1580	1590
GAGTGGTGCAGAGAGAAAAAGAAGCAGTGGAATAGGAGCTTGTCCCTGGGTTCTGGGAGCAGCAGGAA						
GAGTGGTGCAGAGAGAAAAAGAAGCAGTGGAATAGGAGCTTGTCCCTGGGTTCTGGGAGCAGCAGGAA						
7090	7100	7110	7120	7130	7140	7150
7160						
1600	1610	1620	1630	1640	1650	1660
GCACTATGGGCACGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTCAGCAGC						
GCACTATGGGCACGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTCAGCAGC						
7170	7180	7190	7200	7210	7220	7230
1670	1680	1690	1700	1710	1720	1730
AGAACAAATTGCTGAGAGCTATTCAGGCGAACAGCATCTGTTGCAACTCACAGTCTGGGCATCAAGCAGC						
AGAACAAATTGCTGAGAGCTATTCAGGCGAACAGCATCTGTTGCAACTCACAGTCTGGGCATCAAGCAGC						
7240	7250	7260	7270	7280	7290	7300
1750	1760	1770	1780	1790	1800	1810
TCCAGGCAGAAATCCCTGGCTTGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG						
TCCAGGCAGAAATCCCTGGCTTGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG						
7310	7320	7330	7340	7350	7360	7370
1820	1830	1840	1850	1860	1870	1880
GAAAACCTTACATTGACGCGTACGCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA						
GAAAACCTTACATTGACGCGTACGCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA						
7380	7390	7400	7410	7420	7430	7440
1890	1900	1910	1920	1930	1940	1950
ATAACATGACCTGATGAGTACAGAGAAATTAAACAATTACACAAAGCTTAATACATTCTTAATTGAAG						
ATAACATGACCTGATGAGTACAGAGAAATTAAACAATTACACAAAGCTTAATACACTCTTAATTGAAG						
7450	7460	7470	7480	7490	7500	7510
7520						
1960	1970	1980	1990	2000	2010	2020
AATGCGAAACGCGTACGCTTGGAATGCTAGTTGGAGTAATAAATGGGCAAGTTGTGGAAATT						
AATGCGAAACGCGTACGCTTGGAATGCTAGTTGGAGTAATAAATGGGCAAGTTGTGGAAATT						
7530	7540	7550	7560	7570	7580	7590
2030	2040	2050	2060	2070	2080	2090
GGTTTAACATGACCTGATGAGTACAGAGAAATTAAACAATTCTACATAATGATAGTAGGAGGCTTGGTAGGTTAA						
GGTTTAACATGACCTGATGAGTACAGAGAAATTAAACAATTCTACATAATGATAGTAGGAGGCTTGGTAGGTTAA						
7600	7610	7620	7630	7640	7650	7660
2110	2120	2130	2140	2150	2160	2170
GAGTGGTGCAGAGAGAAAAAGAAGCAGTGGAATAGGAGCTTGTCCCTGGGAGCAGCAGCAGC						
GAGTGGTGCAGAGAGAAAAAGAAGCAGTGGAATAGGAGCTTGTCCCTGGGAGCAGCAGCAGC						
7670	7680	7690	7700	7710	7720	7730
2180	2190	2200	2210	2220	2230	2240
ACGCTGCGACGCGTACGCTTGGAATGCTAGGAGGATAGAAGAAGAAGGTTGGAGAGAGAGACAGAGACA						
ACGCTGCGACGCGTACGCTTGGAATGCTAGGAGGATAGAAGAAGAAGGTTGGAGAGAGAGACAGAGACA						
7740	7750	7760	7770	7780	7790	7800
2250	2260	2270	2280	2290	2300	2310
GATGCGAAACGCGTACGCTTGGAATGCTAGGAGGATAGAAGAAGGAGCAGCAGCAGC						
GATGCGAAACGCGTACGCTTGGAATGCTAGGAGGATAGAAGAAGGAGCAGCAGCAGC						

7810	7820	7830	7840	7850	7860	7870	7880
2320	2330	2340	2350	2360	2370	2380	
TACGCGTGTCTTGAGCTTACTCTTGTATTGTAACGAGGATTGTGGAAACTTCTGGGACGCAGGGGGTGGGAA							
TT							
TACGCGTGTCTTGAGCTTACTCTTGTATTGTAACGAGGATTGTGGAAACTTCTGGGACGCAGGGGGTGGGAA							
7890	7900	7910	7920	7930	7940	7950	
2390	2400	2410	2420	2430	2440		
GGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG							
TT							
GGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG							
7960	7970	7980	7990	8000	X		

6. KUNZ-1989-0158-250

HIV-1/HXB3 Human immunodeficiency virus type 1, isolate HXB3,

LOCUS	HIVHXB3	15156 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate HXB3, env region.			
ACCESSTON	M14100			
KEYWORD	+			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate HXB3, proviral DNA.			
ORGANISM	Human immunodeficiency virus type 1			
	Mammal; class: PRIM; enveloped viruses; Retroviridae;			
	Lentivirusinae			
REFERENCE	(+) (base 1 to 15158)			
AUTHORS	Goto, T., Goto, T., Guadron, M., Conroy, R., Schaber, M., Kramer, R., Lazzari, R., Moneti-Goldberg, F. and Reddy, E. P.			
TITLE	Anti-HIV-1 gag gene products synthesized in E. coli are recognized by antibodies present in the sera of AIDS patients			
JOURNAL	Proc. Natl. Acad. Sci. USA 82(14) / 13855			
STANDARD	E. coli (K12) C6000			
FEATURES	from	to/open	description	
pept	{	1	vpr protein, partial (AA at 1)	
pept	1	17	tat protein, exon 2 (first expressed exon)	
pept	17	2634	tat protein, exon 3 (AA at 2633)	
pept	2634	301	rev protein, exon 2 (first expressed exon)	
pept	301	2634	rev protein, exon 3 (AA at 2634)	
pept	301	5643	vpu protein	
pept	5643	6118	envelope polyprotein	
pept	6118	>	nef protein, partial	
pre-msg	{	>	genomic mRNA	
pre-msg	{	>	tat, rev, nef subgenomic mRNA	
pre-msg	{	>	tat, rev, nef subgenomic mRNA intron 1	
IVS	{	>	tat cds intron 2	
IVS	2601	4621	rev cds intron 2	
IVS	4621	5051	rev cds intron 2	
TAT	5051	5571	tat, rev, nef subgenomic mRNA intron 2	
BASE COUNT	1019	a 537 c 736 g 759 t		
ORIGIN	Start site approx. 68 bp upstream of tat initiation codon.			

Initial score = 1874 Optimized Score = 2176 Significance = 0.00

Residue Identity = 89% Matches = 2237 Mismatches = 161

Gaps = 8% Conservative Substitutions = 0

0	10	20	30	40	50	60
ATTTTGTCTTGAGCTTACTCTTGTATTGAGA-TGGGGGTGGAAATGGGGCAC-CATGCTCCTTGG						
TT						
CTTATGAGGAGGAGGAGGAGGAGGAT----GAGAGTGAAGGGAGGAAATATCAGCACTTGTGGAGATGGG						
450	460	470	480	490	500	510

70	80	90	100	110	120
GGGATTCGATG					
ATG					
ATG					
ATG					

2070	2080	2090	2100	2110	2120	2130	
1740	1690	1700	1710	1720	1730	1740	
AATTCTCTTAACTGATTTGAGCCAAUAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAG							
AATCTGCTAACGCGATTTGAGCCAAUAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAG							
2140	2150	2160	2170	2180	2190	2200	
1750	1760	1770	1780	1790	1800	1810	
GCAAGGAATCTTGATTTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAAA							
GGCAAGGAATCTTGATTTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAAA							
2210	2220	2230	2240	2250	2260	2270	2280
1820	1830	1840	1850	1860	1870	1880	1890
CTCATTTCACCACTCTGTGCCTTGGAAATGCTAGTTGGAGTAATAATCTCTGGAACAGATTGGAAATAAC							
CTATTTCACCACTCTGTGCCTTGGAAATGCTAGTTGGAGTAATAATCTCTGGAACAGATTGGAAATAAC							
2290	2300	2310	2320	2330	2340	2350	
1900	1910	1920	1930	1940	1950	1960	
ATGAACTTGATTTGGAAAGGAGAGAAATTACAATTACACAAAGCTTAATACATTCTTAATTGAAGAAATCG							
ACGAACTTGATTTGGAAAGGAGAGAAATTACAATTACACAAAGCTTAATACACTCTTAATTGAAGAAATCG							
2360	2370	2380	2390	2400	2410	2420	
1970	1980	1990	2000	2010	2020	2030	
CAAAACCGGAAAGAAAGAAATTAAACAAAGAATTGGAAATTAGATAATGGCAAGTTGTGGAAATTGGTTT							
CAAAACCGGAAAGAAAGAAATTAAACAAAGAATTGGAAATTAGATAATGGCAAGTTGTGGAAATTGGTTT							
2430	2440	2450	2460	2470	2480	2490	
2040	2050	2060	2070	2080	2090	2100	
AACCTAACAAATTGGGCCTGCTATATAAAATATTCTATAATGATAGTAGGAGGCTTGGTAGGTTAACAAATA							
AACCTAACAAATTGGGCCTGCTATATAAAATTATTCTATAATGATAGTAGGAGGCTTGGTAGGTTAACAAATA							
2500	2510	2520	2530	2540	2550	2560	
2110	2120	2130	2140	2150	2160	2170	
GTTTTGCTGATCTTCTATAGTAAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCCACCTC							
GT1TTTCTCTTCTGATCTGAAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCCACCTC							
2570	2580	2590	2600	2610	2620	2630	2640
2180	2190	2200	2210	2220	2230	2240	2250
CCACGCCCGGAGGAGACCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCC							
CCGATCCCGGAGGAGACCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCC							
2660	2660	2670	2680	2690	2700	2710	
2260	2270	2280	2290	2300	2310	2320	
ATTGATGTTAGTAAACGAGATCTTGGACCTTATCTGGGACGATCTGGGAGCCTTGTGCCTCTTCAGCTACCA							
ATTGATGTTAGTAAACGAGATCTTGGACCTTATCTGGGACGATCTGGGAGCCTTGTGCCTCTTCAGCTACCA							
2720	2730	2740	2750	2760	2770	2780	
2330	2340	2350	2360	2370	2380	2390	
CCGTTGAGGAGACTTACTCTGATTTGTAACGAGATTGTGGAAACTTCTGGGACGCAGGGGGGTGGGAAGCCCT							
CCGTTGAGGAGACTTACTCTGATTTGTAACGAGATTGTGGAAACTTCTGGGACGCAGGGGGGTGGGAAGCCCT							
2790	2800	2810	2820	2830	2840	2850	
2400	2410	2420	2430	X			
CAGATATTCTGTTGAAATCTCTTACAGATATTTGGAGTCAGGAACCTAAAG							
CAGATATTCTGTTGAAATCTCTTACAGATATTTGGAGTCAGGAACCTAAAG							

2860

2870

2880

2890

2900

7. KUNZ-155-PLB3.BEU

HIVNL43 Human immunodeficiency virus type 1, NYS/BRU (LAV-)

LOCUS	HIVNL43	9709 bp ss-RNA	VRL	15-JUN-1989
DEFINITION		Human immunodeficiency virus type 1, NYS/BRU (LAV-1) recombinant clone pNL4-3.		
ACCESSION	M18421			
KEYWORDS	.			
SOURCE		Human immunodeficiency virus type 1 (HIV-1), NYS/BRU (LAV-1) recombinant clone pNL4-3.		
ORGANISM		Human immunodeficiency virus type 1 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.		
REFERENCE	1	(bases 1 to 9709)		
AUTHORS	Buckler,C. E., Buckler-White,A. J., Willey,R. L. and McCoy,J.			
JOURNAL	Unpublished (1988).			
STANDARD	full staff_review			
REFERENCE	2	(bases 1 to 9709)		
AUTHORS	Adechi,A., Bendallman,H. E., Koenig,S., Folks,T., Willey,R., Rabson,A. and Martin,M. A.			
TITLE	Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone			
JOURNAL	J. Virol. 53, 284-291 (1986)			
STANDARD	full staff_review			
REFERENCE	3	(site(s) revisions of [1])		
AUTHORS	Buckler,C. E.			
JOURNAL	Unpublished (1988)			
STANDARD	full staff_review			
COMMENT	Clean copy of sequence [1] kindly provided by Chuck Buckler, NIAID, Bethesda, MD, 24-JUN-1988. The construction of pNL4-3 has been described in [2]. pNL4-3 is a recombinant (infectious) proviral clone that contains DNA from HIV isolates NY5 (5' half) and BRU (3' half). The site of recombination is the EcoRI site at positions 5743-5748. The length and sequence of the vpr coding region corresponds to that of the BRU, SC, SF2, MAL and ELI isolates. The vpr coding region of these isolates is about 18 amino acid residues longer than the vpr coding region of the IIIB isolates. In HIVNL43, this shift is due to a single base deletion (with respect to the IIIB's) at position 5770. The sequence at this position is "atttc" in HIVNL43 and "aitttc" in HIVXB2. The original BRU clone, sequenced by Wain-Hobson, et al. (Cell 40, 9-17 (1985)), and the BRU portion of the pNL4-3 recombinant clone are different clones from the same BRU isolate. Two of the revisions reported in the FEATURES produced changes in amino acid sequences. The revision at position 2421 changes one amino acid residue from 'R' to 'G' in the pol coding region. The revision at positions 8995-9000 changes three amino acid residues from 'AHT' to 'VTP' in the nef coding region.			
FEATURES	from	to/span	description	
pept	790	2292	gag polyprotein	
pept	(2085	5095	pol polyprotein (NH2-terminus uncertain; AA at 2085)	
pept	5041	5619	vif protein	
pept	5558	5849	vpr protein	
pept	5830	6044	tat protein, exon 2 (first expressed exon)	
	8369	8414	tat protein, exon 3 (AA at 8370)	
pept	5568	5044	rev protein, exon 2 (first expressed exon)	
	8369	8643	rev protein, exon 3 (AA at 8371)	
pept	6061	6306	vpu protein	
pept	6221	8785	envelope polyprotein	
pept	8787	9107	nef protein	

pre-msg	456	9626	genomic mRNA
pre-msg	455	9626	tat, rev, nef subgenomic mRNA
IVS	744	8776	tat, rev, nef mRNA intron 1
IVS	6045	6368	tat cds intron 2
IVS	6045	8368	rev cds intron 2
IVS	6045	8368	tat, rev, nef mRNA intron 2
LTR	1	634	5' LTR
LTR	9076	9709	3' LTR
rpt	454	550	R repeat 5' copy
rpt	7529	9626	R repeat 3' copy
binding	377	386	Spi binding site III
binding	388	397	Spi binding site II
binding	399	408	Spi binding site I
binding	636	653	primer (Lys-tRNA) binding site
site	5743	5748	EcoRI site of recombination
recomb	5743	5744	HIV-1 isolate NYS DNA end/HIV-1 isolate LAV DNA start
revision	182	183	at in [3]; tg in [1]
revision	194	194	g in [3]; c in [1]
revision	2421	2421	g in [3]; a in [1]
revision	2895	5000	tcacac in [3]; ctcaca in [1]
revision	9415	9415	c in [3]; a in [1]
signal	9602	9607	mRNA polyadenylation signal
BASE COUNT	3421	a 1750 c 2366 g 2166 t	
ORIGIN	5'	terminus of NYS LTR	

Initial Score = 1723 Optimized Score = 2169 Significance = 0.00
Residue Identity = 89% Matches = 2226 Mismatches = 179
Gaps = 84 Conservative Substitutions = 0

X	10	20	30	40	50	60	
ATGAGAGTGAAGGAGAAATATCAGICA-CTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCT-TGG-G							
TTGATAGAICATATACTAAAGAGGAGAAAGACAGTGCGCAATGAGAGTGAAGGAGAAGTATCA-GCACTTGTGGAG							
X	6190	6200	6210	6220	6230	6240	6250

70	80	90	100	110	120	
ATATTGAT-GATCTGTAG-----TGCTAC-AGAAAAAATTGTGGGTC-ACAGT-CT-----ATTATGGGG						
: : : : : : : : : : : : : : : : : : :						
ATGAGGGCTGGAAATGCGGCCACCATRCCCTTGGATATTGATGATCTGTAGTGCTACAGAAAAATTGT-GGG						
6260	6270	6280	6290	6300	6310	6320

```

  130      140      150      160      170      180      190
TACCTGTGTTGGAAAGCAACCACACTCTATTTTG-TGCATCAGATGCTAAAGCATATGATACAGA-GGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCACAGTCTATTATGCGG----TACCTGTGTTGGAAAGCAACCACACTCTATTTTG-TGCATCAGATGCT
6330      6340      6350      6360      6370      6380      6390

```

260 270 280 290 300 310 320
 GTAAATGTGACAGAAAAATTAAACATGTG---GAAAA--ATGACATGGTAGAACAGATG-CATGAGGGATATA
 :
 CAGAAGT---AT-TATTTGGTAA-ATGTGACAGAAAAATTAAACAT-GTGGAA-AAATGACAT---GGTAGA
 6470 6480 6490 6500 6510 6520

330 340 350 360 370 380 390
 ATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-TAAAA--TTAACCCCCACTCTGTGTTAGTTAA-AG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 A-CAGATGCGATGAGGATATAA---TCAGTTATGGGATCAAAGCCTAAAGCCA---TGTGTAAAATTAAACCC
 6530 6540 6550 6560 6570 6580 6590

400 410 420 430 440 450
TGCACGTGATTTGG----GAAATGCTACTAAT---ACCAATACTAGTAATAACCAATAGTAGTAGCGGGGAAATG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CACTCTGTATTGTTAAAGTGCACTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGGGAGAATG
6500 6610 6620 6630 6640 6650 6660

460 470 480 490 500 510 520
ATGATGGGAAAGGGAGAGATAAAAAACTGCTCTTCATAATATCAGCACAGNATAAGAGGTAAAGGTGCAGAAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATAATGGAGAAAGGGAGAGATAAAAAACTGCTCTTCATAATATCAGCACAGCATAAGAGATAAGGTGCAGAAA
6670 6680 6690 6700 6710 6720 6730

530 540 550 560 570 580 590 600
GAATATGCATTTTATAAACTGATATAACCAATAGATAATGATACTACAGCTATACGTTGACAAGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAATATGCATTCTTTATAAACTGATATAGTACCAATAGATAA-----TACCAAGCTATAGGTTGATAAGT
6740 6750 6760 6770 6780 6790

610 620 630 640 650 660 670
TGTAACACUTGCTAGTCATTACACAGGCCCTGTCAGGATCCTTGAGCCAATTCCCATACATTATTGTGCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGTAACACUTGCTAGTCATTACACAGGCCCTGTCAGGATCCTTGAGCCAATTCCCATACATTATTGTGCC
6800 6810 6820 6830 6840 6850 6860 6870

680 690 700 710 720 730 740
CCGGCTGGTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCGGCTGGTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGC
6880 6890 6900 6910 6920 6930 6940

750 760 770 780 790 800 810
ACAGTACAACTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACAGTACAAATGTAACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAA
6950 6960 6970 6980 6990 7000 7010

820 830 840 850 860 870 880
GAAGAGGTAGTAATTAGATCTGCAATTTCACAGACAATGCTAAAACCATAATAGTACAGCTGAACCAAATCT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAAGAGGTAGTAATTAGATCTGCAATTTCACAGACAATGCTAAAACCATAATAGTACAGCTGAACACATCT
7020 7030 7040 7050 7060 7070 7080

890 900 910 920 930 940 950 960
GTTAAATTAATTTACAAAGACCCAAACAAACAATACAAGAAAAAGTATCOGTATCCAGAGGGGGACCAGGGAGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTAAACATTAAATTACAAACACCCAAACAAACAATACAAGAAAAAGTATCCGTATCCAGAGGGGGACCAGGGAGA
7090 7100 7110 7120 7130 7140 7150

970 980 990 1000 1010 1020 1030
GCATTTGTTACAAATGGAAAAATAAGGACAAGCACATTGTAACATTAGTAGAGCAAAATGCAAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCATTTGTTACAAATGGAAAAATAAGGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAAT
7160 7170 7180 7190 7200 7210 7220 7230

1040 1050 1060 1070 1080 1090 1100
GCCACTTTAAACAGATAGCTAGCAAATTAAAGAGAACAAATTGGAAATAATAAAACAATAATCTTAAGCAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GUCACTTTAAACAGAATAGGACAAATTGAGAACAAATTGGAAATAATAAAACAATAATCTTAAGCAA
7240 7250 7260 7270 7280 7290 7300

1110 1120 1130 1140 1150 1160 1170
TCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTAATTGTGGAGGGGAATTTCCTACTGTAATTCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTAATTGTGGAGGGGAATTTCCTACTGTAATTCA
7310 7320 7330 7340 7350 7360 7370

1150 1150 1200 1210 1220 1230 1240
ACACAACCTGTTAATAGTACTTGGTTAACAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGT
1150 1150 1200 1210 1220 1230 1240
ACACAACCTGTTAATAGTACTTGGTTAACAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGT
7330 7330 7400 7410 7420 7430 7440

1250 1250 1270 1280 1290 1300 1310 1320
GACACAATCACACTCCCATCGAAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTAT
1250 1250 1270 1280 1290 1300 1310 1320
GACACAATCACACTCCCATCGAAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTAT
7450 7450 7470 7490 7490 7500 7510

1330 1340 1350 1360 1370 1380 1390
GCCGCTCCCATCGACGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAAGAGATGGTGTT
1330 1340 1350 1360 1370 1380 1390
GCCGCTCCCATCGACGACAAATTAGATGTTCATCAAATATTACTGGGCTGCTATTAAACAAGAGATGGTGTT
7520 7530 7540 7550 7560 7570 7580 7590

1400 1410 1420 1430 1440 1450 1460
AATAACAACAAATGGG1CCGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAATTGGAGAAGTGAATTA
1400 1410 1420 1430 1440 1450 1460
AATAACAACAAATGGG1CCGAGATCTTCAGACCTGGAGGGAGATATGAGGGACAATTGGAGAAGTGAATTA
7600 7610 7620 7630 7640 7650 7660

1470 1480 1490 1500 1510 1520 1530
TATAAATATAAAGTAGTAAAGAATTGAAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAAG
1470 1480 1490 1500 1510 1520 1530
TATAAATATAAAGTAGTAAAGAATTGAAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAAG
7670 7680 7690 7700 7710 7720 7730

1540 1550 1560 1570 1580 1590 1600
AGAGAAAAAGAGGCACTGGGAATAGGAGCTTGTCTGGTTCTGGGAGCAGCAGGAAGCACTATGGGC
1540 1550 1560 1570 1580 1590 1600
AGAGAAAAAGAGGCACTGGGAATAGGAGCTTGTCTGGTTCTGGGAGCAGCAGGAAGCACTATGGGC
7740 7750 7760 7770 7780 7790 7800

1610 1620 1630 1640 1650 1660 1670
-GCACGGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAAATT
1610 1620 1630 1640 1650 1660 1670
TGCAC-GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGAACAAATT
7810 7820 7830 7840 7850 7860 7870

1680 1690 1700 1710 1720 1730 1740 1750
GCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAG
1680 1690 1700 1710 1720 1730 1740 1750
GCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAACAGCTCCAGGCAAG
7880 7890 7900 7910 7920 7930 7940 7950

1760 1770 1780 1790 1800 1810 1820
AATCCTGGCTG1GGAAAGATACCTAAAGGATCAACAGCTCTGGGGATTGGGGTTGCTCTGGAAAACCTCAT
1760 1770 1780 1790 1800 1810 1820
AATCCTGGCTG1GGAAAGATACCTAAAGGATCAACAGCTCTGGGGATTGGGGTTGCTCTGGAAAACCTCAT
7960 7970 7980 7990 8000 8010 8020

1830 1840 1850 1860 1870 1880 1890
TTGCACCACTGCTGTCCTTGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGAATAACATGAC
1830 1840 1850 1860 1870 1880 1890
TTGCACCACTGCTGTCCTTGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTGGAATAACATGAC
8030 8040 8050 8060 8070 8080 8090

1900 1910 1920 1930 1940 1950 1960
CTGGATGGAGTGGACAGAGAAATTAAACAATTACACACAAGCTTAATACATTCCCTTAATTGAAGAATCGCAAAA
1900 1910 1920 1930 1940 1950 1960
CTGGATGGAGTGGACAGAGAAATTAAACAATTACACACAAGCTTAATACACTCCCTTAATTGAAGAATCGCAAAA
8100 8110 8120 8130 8140 8150 8160

1970	1980	1990	2000	2010	2020	2030
CCAGCAAGAAAGAATGAAACAGAATTATTGGAATTAGATAAATGGGCAAGTTGTGGAATTGGTTAACAT 8170	8190	8190	8200	8210	8220	8230
2040	2050	2060	2070	2080	2090	2100
AACAAATTGGCTTGTTATATAAAAATATTCTATAATGATAGTAGGAGGCTTGGTAGGTTAACAT 8240	8250	8260	8270	8280	8290	8300
2110	2120	2130	2140	2150	2160	2170
TGCCTACTTCTATAAGTGAATAGTATTAGGCAGGGATATTCAACCATTATCGTTTCAGACCCACCTCCCAAC 8320	8330	8340	8350	8360	8370	8380
2180	2190	2200	2210	2220	2230	2240
CCCGAGGGGACGCCACAGGCCAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGACAGATCCATTG 8380	8400	8410	8420	8430	8440	8450
2250	2260	2270	2280	2290	2300	2310
ATTAGTGAACGCGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCAACCGCT 8460	8470	8480	8490	8500	8510	8520
2320	2330	2340	2350	2360	2370	2380
TGAGAGACTTACTCTGATTGTAACGAGGATTGTGGAACCTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAAT 8530	8540	8550	8560	8570	8580	8590
2390	2400	2410	2420	2430	2440	
ATTCGGAAATCTCCTACAGTATTGGAGTCAGGAACCTAAAG 8600	8610	8620	8630	X		

8. KUNZ-158-CL33, SEE

HIVHXB2CG Human immunodeficiency virus type 1 (HXB2), complete

LOCUS HIVHXB2CG 9718 bp ss-RNA VRL 25-SEP-1987
 DEFINITION Human immunodeficiency virus type 1 (HXB2), complete genome; HIV1/HTLV-III/LAV reference genome.
 ACCESSION K03455
 KEYWORDS TAR protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activator.
 SOURCE HTLV-III/LAV (isolate HXB2) proviral DNA.
 ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
 REFERENCE 1 (bases 1 to 653)
 AUTHORS Starcich,B., Rother,L., Josephs,S. F., Okamoto,T., Gallo,R. C. and Wong-Staal,F.
 TITLE Characterization of long terminal repeat sequences of HTLV-III
 JOURNAL Science 227, 538-540 (1985)
 STANDARD full staff_review
 REFERENCE 2 (bases 493 to 674; 9577 to 9718)
 AUTHORS Rathner,L., Haseltine,W., Patarca,R., Livak,K. J., Starcich,B., Josephs,S. J., Doreen,E. R., Rafalski,J. A., Whitehorn,E. A., Baumleiter,K., Tammiff,L., Petteway,S. R. Jr., Pearson,M. L.,

Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. and Wong-Staal, F.
TITLE
JOURNAL Nature 313, 277-284 (1985)
STANDARD full staff_review
REFERENCE 3 (bases 8761 to 9060)
AUTHORS Fisher, A. G., Ratner, L., Mitsuya, H., Marseille, L. M., Harper, M. E., Broder, S., Gallo, R. C. and Wong-Staal, F.

TITLE Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects
JOURNAL Science 233, 655-659 (1986)
STANDARD full staff_review
REFERENCE 4 (bases 1 to 9635)
AUTHORS Ratner, L., Fisher, A., Jagodzinski, L. L., Mitsuya, H., Liou, R.-S., Gallo, R. C. and Wong-Staal, F.

TITLE Complete nucleotide sequences of functional clones of the AIDS virus
JOURNAL AIDS Res. Hum. Retroviruses 3, 57-69 (1987)
STANDARD full staff_review
REFERENCE 5 (sites; tat mRNA and other transcript boundaries)
AUTHORS Arya, S. K., Guo, C., Josephs, S. F. and Wong-Staal, F.

TITLE Trans-activator gene of human T-lymphotropic virus type III (HTLV-III)
JOURNAL Science 229, 69-73 (1985)
STANDARD full staff_review
REFERENCE 6 (sites; tat mRNA)
AUTHORS Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. and Haseltine, W.

TITLE Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III
JOURNAL Science 229, 74-77 (1985)
STANDARD full staff_review
REFERENCE 7 (sites; mRNA splice sites)
AUTHORS Rabson, A. B., Daugherty, D. F., Venkatesan, S., Boulukos, K. E., Barin, S. I., Folks, T. M., Feorino, P. and Martin, M.

TITLE Transcription of novel open reading frames of AIDS retrovirus during infection of lymphocytes
JOURNAL Science 228, 1388-1390 (1985)
STANDARD full staff_review
REFERENCE 8 (sites; 27K antigen cds)
AUTHORS Allan, J. S., Coligan, J. E., Lee, T.-H., McLane, M. F., Kanki, P. J., Groopman, J. E. and Essex, M.

TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients
JOURNAL Science 230, 810-813 (1985)
STANDARD full staff_review
REFERENCE 9 (sites; gp160 and gp120 coding sequences)
AUTHORS Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H. and Essex, M.

TITLE Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III
JOURNAL Science 228, 1091-1094 (1985)
STANDARD full staff_review
REFERENCE 10 (sites; regulatory sequences in the LTR)
AUTHORS Rosen, C. A., Sodroski, J. G. and Haseltine, W. A.

TITLE The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat
JOURNAL Cell 41, 813-823 (1985)
STANDARD full staff_review
REFERENCE 11 (review; bases 1 to 8718)
AUTHORS Van Beveren, C., Coffin, J. and Hughes, S.

TITLE Appendix B: HTLV-3 genome
JOURNAL (Eds.) Weiss, R., Teich, N., Varmus, H. and Coffin, J. (Eds.); RNA Tumor Viruses, Second Edition, 2: 1102-1123 and 1147-1148; Cold Spring Harbor Laboratory, New York (1985)

STANDARD full staff_review
REFERENCE 12 (sites; trans-activator function and TAR sequence)

AUTHORS Rosen,C. A. , Sodroski,J. G. , Goh,W. C. , Dayton,A. I. , Lippke,J. and Haseltine,W. A.
TITLE Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III
JOURNAL Nature 319, 555-559 (1986)
STANDARD full staff_review
REFERENCE 13 (sites; pol coding sequence)
AUTHORS di Marzo Veronese,F. , Copeland,T. D. , DeVico,A. L. , Rahman,R. , Oroszlan,S. , Gallo,R. C. and Sarngadharan,M. G.
TITLE Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV
JOURNAL Science 231, 1289-1291 (1986)
STANDARD full staff_review
REFERENCE 14 (sites; 23K sor gene product)
AUTHORS Kan,N. C. , Franchini,G. , Wong-Staal,F. , DuBois,G. C. , Robey,W. G. , Lautenberger,J. A. and Papas,T. S.
TITLE Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera
JOURNAL Science 231, 1553-1555 (1986)
STANDARD full staff_review
REFERENCE 15 (sites; pol NH2-terminal region)
AUTHORS Kramer,R. A. , Schaber,M. D. , Skalka,A. M. , Ganguly,K. , Wong-Staal,F. and Reddy,E. P.
TITLE HTLV-III gag protein is processed in yeast cells by the virus pol-protease
JOURNAL Science 231, 1580-1584 (1986)
STANDARD full staff_review
REFERENCE 16 (sites; sor 23K protein)
AUTHORS Lee,T.-H. , Coligan,J. E. , Allan,J. S. , McLane,M. F. , Groopman,J. E. and Essex,M.
TITLE A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses
JOURNAL Science 231, 1546-1549 (1986)
STANDARD full staff_review
REFERENCE 17 (sites; sor 23K protein)
AUTHORS Sodroski,J. , Goh,W. C. , Rosen,C. , Tartar,A. , Portetelle,D. , Bunn,A. and Haseltine,W.
TITLE Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions
JOURNAL Science 231, 1549-1553 (1986)
STANDARD full staff_review
REFERENCE 18 (sites; Sp1 binding sites in the promoter region)
AUTHORS Jones,K. A. , Kadonaga,J. T. , Luciw,P. A. and Tjian,R.
TITLE Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1
JOURNAL Science 232, 755-759 (1986)
STANDARD full staff_review
REFERENCE 19 (sites; acceptor and donor splice sites for tat and 27K)
AUTHORS Arya,S. K. and Gallo,R. C.
TITLE Three novel genes of human T-lymphotropic virus type III: Immune reactivity of their products with sera from acquired immune deficiency syndrome patients
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 2209-2213 (1986)
STANDARD full staff_review
REFERENCE 20 (sites; deletion mutants in the tat gene)
AUTHORS Dayton,A. I. , Sodroski,J. G. , Rosen,C. A. , Goh,W. C. and Haseltine,W. A.
TITLE The trans-activator gene of the human T cell lymphotropic virus type III is required for replication
JOURNAL Cell 44, 941-947 (1986)
STANDARD full staff_review
REFERENCE 21 (sites; env gene conserved/varable regions; separate entries)
AUTHORS Willey,R. , Rutledge,R. A. , Dias,S. , Folks,T. , Theodore,T. , Buckler,C. E. and Martin,M. A.
TITLE Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 5038-5042 (1986)

STANDARD full staff_review
REFERENCE 22 (sites; trs cds boundaries)
AUTHORS Sodroski,J. , Goh,W. C. , Rosen,C. , Dayton,A. I. , Terwilliger,E. and Haseltine,W.

TITLE A second post-transcriptional trans-activator gene required for HTLV-III replication
JOURNAL Nature 321, 412-417 (1986)

STANDARD full staff_review
REFERENCE 23 (sites; trs cds boundaries)
AUTHORS Feinberg,M. B. , Jarret,R. F. , Aldovini,A. , Gallo,R. C. and Wong-Staal,F.

TITLE HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA
JOURNAL Cell 46, 807-817 (1986)

STANDARD full staff_review
REFERENCE 24 (sites; env gene conserved/variable regions; separate entries)
AUTHORS Starcich,B. R. , Hahn,B. H. , Shaw,G. M. , McNeely,P. D. , Modrow,S. , Wolf,H. , Parks,E. S. , Parks,W. P. , Josephs,S. F. , Gallo,R. C. and Wong-Staal,F.

TITLE Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS
JOURNAL Cell 45, 637-648 (1986)

STANDARD full staff_review
REFERENCE 25 (sites; tar or transactivator target)
AUTHORS Wright,C. M. , Felber,B. K. , Paskalis,H. and Pavlakis,G. N.

TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus
JOURNAL Science 234, 988-992 (1986)

STANDARD full staff_review
REFERENCE 26 (sites; 3' orf mutations)
AUTHORS Tervilliger,E. , Sodroski,J. G. , Rosen,C. A. and Haseltine,W. A.

TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity
J. Virol. 60, 754-760 (1986)

STANDARD full staff_review
REFERENCE 27 (sites; pol p34 terminus)
AUTHORS Lightfoot,M. M. , Coligan,J. E. , Folks,T. M. , Fauci,A. S. , Martin,M. A. and Venkatesan,S.

TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus
J. Virol. 60, 771-775 (1986)

STANDARD full staff_review
REFERENCE 28 (sites; promoter, TAR, tat-III mutants)
AUTHORS Muesing,M. A. , Smith,D. H. and Capon,D. J.

TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein
JOURNAL Cell 48, 681-701 (1987)

STANDARD full staff_review
REFERENCE 29 (sites; envelope protein epitopes)
AUTHORS Modrow,S. , Hahn,B. H. , Shaw,G. M. , Gallo,R. C. , Wong-Staal,F. and Wolf,H.

TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions
J. Virol. 61, 570-578 (1987)

STANDARD full staff_review
REFERENCE 30 (sites; trs/ert protein)
AUTHORS Goh,W. C. , Sodroski,J. G. , Rosen,C. A. and Haseltine,W. A.

TITLE Expression of the ert gene protein of human T-lymphotropic virus type III (HTLV-III/LAV) in bacteria
J. Virol. 61, 633-637 (1987)

STANDARD full staff_review
REFERENCE 31 (sites; inducible enhancer element)

AUTHORS	Nebel,B. and Baltimore,D.		
TITLE	An inducible transcription factor activates expression of human immunodeficiency virus in T cells		
JOURNAL	Nature 326, 711-713 (1987)		
STANDARD	full staff_review		
REFERENCE	32 (bases 5611 to 5611; revises [4])		
AUTHORS	Ratner,L.		
JOURNAL	Unpublished (1987) Washington U Med School, St. Louis, MO		
STANDARD	full staff_review		
REFERENCE	33 (sites; long terminal repeat)		
AUTHORS	Paterca,R., Heath,C., Goldenberg,G. J., Rosen,C. A., Sodroski,J. G., Haseltine,W. A. and Hansen,U. M.		
TITLE	Transcription directed by the HIV long terminal repeat in vitro		
JOURNAL	AIDS Res. Hum. Retroviruses 3, 41-55 (1987)		
STANDARD	full staff_review		
REFERENCE	34 (sites; R orf)		
AUTHORS	Wong-Staal,F., Chanda,P. K. and Ghrayeb,J.		
TITLE	Human immunodeficiency virus: the eighth gene		
JOURNAL	AIDS Res. Hum. Retroviruses 3, 33-39 (1987)		
STANDARD	full staff_review		
REFERENCE	35 (sites; sor)		
AUTHORS	Fisher,A. G., Ensoli,B., Ivanoff,L., Chamberlain,M., Petteway,S., Ratner,L., Gallo,R. C. and Wong-Staal,F.		
TITLE	The sor gene of hiv-1 is required for efficient virus transmission in vitro		
JOURNAL	Science 237, 888-893 (1987)		
STANDARD	full staff_review		
COMMENT	Sequence for [3] kindly provided in computer-readable form by L. Ratner, 19-AUG-1986.		
The HXB2 sequence is being used as a reference genome for all the HIV entries because it has been derived from a demonstrably infectious clone. Hence not all of the "sites" references above were concerned with this isolate.			
FEATURES	from	to/span	description
pept	788	2291	gag polyprotein
pept	/	2357	pol polyprotein (NH2-terminus uncertain; AA at 2357)
pept	5040	5618	sor 23K protein
pept	5558	5794	R (ORF) protein
pept	5830	6044	tat protein, exon 2 (first expressed exon)
	5378	8423	tat protein, exon 3
pept	5968	6044	trs protein, exon 2 (first expressed exon)
	8379	8652	trs protein, exon 3
pept	6224	8784	envelope polyprotein
pept	6736	9167	27K protein (premature termination)
mRNA	455	5635	HXB2 genomic mRNA
pre-msg	455	5635	tat, trs, 27K subgenomic mRNA
IVS	6045	8377	tat intron 1
IVS	6045	8377	trs intron 2
IVS	6045	8377	27K mRNA intron 2
IVS	743	5776	tat, trs, 27K mRNA intron 1
IVS	6045	8377	tat, trs intron 2
LTR	1	634	5' LTR
LTR	9085	9718	3' LTR
rpt	454	551	R repeat 5' copy
rpt	9538	9635	R repeat 3' copy
binding	377	386	Spl binding site III
binding	383	397	Spl binding site II
binding	392	406	Spl binding site I
binding	636	653	primer (Lys-tRNA) binding site
revision	5611	5611	g in [32]; a in [4]
signal	5611	5616	HXB2 mRNA polyadenylation signal
BASE COUNT	3411	e 1773 c 2370 g 2164 t	
ORIGIN	425 bp upstream of PvuII site; 5' end of proviral genome.		

Initial Score = 1858, Optimized Score = 2164, Significance = 0.00

Residue Identity = 89% Matches = 2231 Mismatches = 162
Gaps = 100 Conservative Substitutions = 0

X 10 20 30 40 50 60
ATGAGAGTAA-AGRAAAGA-ATATCAGCACTTGAGAGA-TGGGGGGTGGAAATGGGGCAC-CATGCTCCTTGG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTATAGAAGAGACAGAAGACAGTGGCAAT---GAGAGTGAAGGAGAAAATACAGCACTTGTGGAGATGGG
6200 6210 6220 6230 6240 6250 6260

70 80 90 100 110 120
GATATTGATA---AT-CT-GTAGTGCACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GGTGGAGATEGGCACCCATCTCCCTGGATGTTGATGATCTGTAGTGCACAGAAAAATTGTGGGTACAG
6270 6280 6290 6300 6310 6320 6330

130 140 150 160 170 180
-CT---GTGTCUAA---GGAAAGCAA-CCACCA-CTCTATTGTGATCAGATGCTAAAGCATATGAT
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TCTATATGGGTAACCTGTGTGGAAAGCAACCAACCCTCTATTG-----TGC---ATCAGATGCT
6340 6350 6360 6370 6380 6390

190 200 210 220 230 240
A---CAGAAGG-TACATA---AT---GTTTGGGCCACACATGCCTG---TGTACCCACAGA-CCCCAACCCAC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAACCATATGATACAGAGGTACATAATTTGGGC-CACA---CATGCCTGTGACCCACAGACCCCAACC-C
6400 6410 6420 6430 6440 6450 6460

250 260 270 280 290 300 310
AAGAAGTAGTATTGCTAAATGTCACAGAAAATTAAACATGTGGAAA---ATGACATGGTAGAACAGATG-C
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
A-CAYGAAGHTA---GTA-TTG---GTAAAT-GTGACAC---GAAAATTTGACAT-GTGGAA-AAATGAC
6470 6480 6490 6500 6510 6520

320 330 340 350 360 370 380
ATGAGGGATAATACTCGT-TTTATG-GGATCAAAGCCTAAAGCCATGTG-TAAAA---TTAACCCCACACTGTG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AT---GCTAGAA-CAGATGCATGAGGATATAA---TCAGTTTATGGGATCAAAGCCTAAAGCCA---TGTGT
6530 6540 6550 6560 6570 6580

390 400 410 420 430 440
TAAUTTAA-ACTGCACTGATTGG---GGAATGCTACTAAT---ACCAACTAGTAATACCAATAGTAGTA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAAATTAAACCCCACTCTGTGTTAATTTAAAGTGC-ACTGATTGAGAATGATACTAATACCAATAGTAGTA
6590 6600 6610 6620 6630 6640 6650

450 460 470 480 490 500 510
GCGAAGGAAAATGATGATGGAGAAGAGAGATAAAAAACTGCTCTTCAATATCAGCACAGNATAAGAGGTA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCGGAGAGAATGATAATGGAGAAGAGAGATAAAAAACTGCTCTTCAATATCAGCACAGCATAAGAGGTA
6660 6670 6680 6690 6700 6710 6720

520 530 540 550 560 570 580
AGGTGCAAGAAAGAATATGCAATTITTTATAAACTTGATATAATACCAATAGATAATGATACTACCGCTATA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AGGTGCAAGAAAGAATATGCAATTITTTATAAACTTGATATAATACCAATAGATAATGATACTACCGCTATA
6730 6740 6750 6760 6770 6780 6790

590 600 610 620 630 640 650 660
CGTTGACAAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCACAAAGGTATCCTTGAGCCAATTCCCATAC
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCTTGACAAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCACAAAGGTATCCTTGAGCCAATTCCCATAC
6800 6810 6820 6830 6840 6850 6860

670 680 690 700 710 720 730
ATTATTTGCTGGGGGGGGCTGGTTGGATTCTAAATGTAATAAGACGTTCAATGGAACAGGACCATGTA
||| ||| ||| ||| ||| ||| ||| ||| ||| |||

ATTATTGTGCCCCGGCTGGT TTTGCATTCATAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
6870 6880 6890 6900 6910 6920 6930 6940
740 750 760 770 780 790 800
CAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA
6950 6960 6970 6980 6990 7000 7010
810 820 830 840 850 860 870
GTCTAGCAAGAAGAAGGTAGTAATTAGATCTGCCAATTACACAGACAATGCTAAAACCATAATAGTACAGC
||||| ||||| ||||| ||||| ||||| ||||| |||||
GTCTAGCAAGAAGAAGGTAGTAATTAGATCTGTCATTACACGGACAATGCTAAAACCATAATAGTACAGC
7020 7030 7040 7050 7060 7070 7080
880 890 900 910 920 930 940
TGAACCAAATCTGAAATTAAATTGTACAAGACCCAAACAACAAATACAAGAAAAAGTATCCGTATCCAGAGGG
||||| ||||| ||||| ||||| ||||| ||||| |||||
TGAACACATCTGAAATTAAATTGTACAAGACCCAAACAACAAATACAAGAAAAAGAATCCGTATCCAGAGAG
7090 7100 7110 7120 7130 7140 7150
950 960 970 980 990 1000 1010 1020
GACCAGGGAGAACATTGTACAATAGGAAAAATAGGAAATATGAGACAAAGCACATTGTAACATTAGTAGAG
||||| ||||| ||||| ||||| ||||| |||||
GACCAGGGAGAACATTGTACAATAGGAAAAATAGGAAATATGAGACAAAGCACATTGTAACATTAGTAGAG
7160 7170 7180 7190 7200 7210 7220
1030 1040 1050 1060 1070 1080 1090
CAAAATGCAATTGCACTTTAAACAGATAGCTAGCAAATTAAAGAGAACAAATTGGAAATAATAAAACAATAA
||||| ||||| ||||| ||||| ||||| |||||
CAAATGGAATAACACTTTAAACAGATAGCTAGCAAATTAAAGAGAACAAATTGGAAATAATAAAACAATAA
7230 7240 7250 7260 7270 7280 7290 7300
1100 1110 1120 1130 1140 1150 1160
TCTTTAAGCAATCTCAGGGGGACCCAGAAATTGTAAACGCACAGTTTAATTGTGGAGGGAAATTTTCT
||||| ||||| ||||| ||||| ||||| |||||
TCTTTAAGCAATCTCAGGGGGACCCAGAAATTGTAAACGCACAGTTTAATTGTGGAGGGAAATTTTCT
7310 7320 7330 7340 7350 7360 7370
1170 1180 1190 1200 1210 1220 1230
ACTGTAATTCAACACACAACGTGTTAATAGTACTTGTTTAATAGTACTTGGAGTAGTGAAGGGTCAAATAACA
||||| ||||| ||||| ||||| ||||| |||||
ACTGTAATTCAACACACAACGTGTTAATAGTACTTGTTTAATAGTACTTGGAGTAGTGAAGGGTCAAATAACA
7380 7390 7400 7410 7420 7430 7440
1240 1250 1260 1270 1280 1290 1300
CTGAYAGGAAGTACACAAATCACACTCCCATGCAGAACAAATTATAAACATGTGGCAGGAAGTAGGAA
||||| ||||| ||||| ||||| ||||| |||||
CTGAYAGGAAGTACACAAATCACCTCCCATGCAGAACAAATTATAAACATGTGGCAGAACAGTAGGAA
7450 7460 7470 7480 7490 7500 7510
1310 1320 1330 1340 1350 1360 1370 1380
AAGCAATGTATGCCCTCCCATCAGCGAACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAA
||||| ||||| ||||| ||||| ||||| |||||
AAGCAATGTATGCCCTCCCATCAGTGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAA
7520 7530 7540 7550 7560 7570 7580
1390 1400 1410 1420 1430 1440 1450
GAGATGGTGTAAATAACAAACAAATGGTCCGAGATCTTCAGACCTGGAGGGAGAGATGAGGGACAAATTGGA
||||| ||||| ||||| ||||| ||||| |||||
GAGATGGTGTAAATAACAAACAAATGGTCCGAGATCTTCAGACCTGGAGGGAGAGATGAGGGACAAATTGGA
7590 7600 7610 7620 7630 7640 7650 7660
1460 1470 1480 1490 1500 1510 1520
GAGATGGTGTAAATAACAAACAAATGGTAAACCAATTAGGGAGTAGCACCCACCAAGGGCAAAGAGAA
||||| ||||| ||||| |||||

GAATCTGAAATTTATAAATAT/AAAGTAGTAAAATTCACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
7670 7680 7690 7700 7710 7720 7730
1530 1540 1550 1560 1570 1580 1590
GAATCTGAAATTTATAAATTCACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
1530 1540 1550 1560 1570 1580 1590
GAATCTGAAATTTATAAATTCACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
7740 7750 7760 7770 7780 7790 7800
1600 1610 1620 1630 1640 1650 1660
GCACATATGGGCCAACGCTCAATGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGAGCAGCAGC
1600 1610 1620 1630 1640 1650 1660
GCACATATGGGCCAACGCTCAATGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGAGCAGCAGC
7810 7820 7830 7840 7850 7860 7870
1670 1680 1690 1700 1710 1720 1730 1740
AGAACAAATTTCGCTAGGGCTATTGAGGCCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
1670 1680 1690 1700 1710 1720 1730 1740
AGAACAAATTTCGCTAGGGCTATTGAGGCCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
7880 7890 7900 7910 7920 7930 7940
1750 1760 1770 1780 1790 1800 1810
TCCGGCAAGAACTCTGGCTCTGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG
1750 1760 1770 1780 1790 1800 1810
TCCGGCAAGAACTCTGGCTCTGGAAAGATAACCTAAAGGATCAACAGCTCCTAGGGATTGGGGTTGCTCTG
7950 7960 7970 7980 7990 8000 8010 8020
1820 1830 1840 1850 1860 1870 1880
GAAACTCATTTGACCACTGCTGTGCGCTTGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTTGGA
1820 1830 1840 1850 1860 1870 1880
GAAACTCATTTGACCACTGCTGTGCGCTTGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATCTGGG
8030 8040 8050 8060 8070 8080 8090
1890 1900 1910 1920 1930 1940 1950
ATAACATGACTTGGATGGAGTGGGACAGAGAAATTAAACAAATTACACAAAGCTTAATACATTCTTAATTGAAG
1890 1900 1910 1920 1930 1940 1950
ATAACACGGACTTGGATGGAGTGGGACAGAGAAATTAAACAAATTACACAAAGCTTAATACACTCTTAATTGAAG
8100 8110 8120 8130 8140 8150 8160
1960 1970 1980 1990 2000 2010 2020
AATGGCAAAACGCAAGAAAGAAATTAAACAAAGAAATTGGAAATTAGATAAAATGGGCAAGTTGTGGAAATT
1960 1970 1980 1990 2000 2010 2020
AATGGCAAAACGCAAGAAAGAAATTGGAAATTAGATAAAATGGGCAAGTTGTGGAAATT
8170 8180 8190 8200 8210 8220 8230
2030 2040 2050 2060 2070 2080 2090 2100
GGTTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATAATGATAGTAGGAGGCTTGGTAGGTTAA
2030 2040 2050 2060 2070 2080 2090 2100
GGTTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATAATGATAGTAGGAGGCTTGGTAGGTTAA
8240 8250 8260 8270 8280 8290 8300
2110 2120 2130 2140 2150 2160 2170
GAATAGTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACTTATCGTTTCAGACCC
2110 2120 2130 2140 2150 2160 2170
GAATAGTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACTTATCGTTTCAGACCC
8310 8320 8330 8340 8350 8360 8370 8380
2180 2190 2200 2210 2220 2230 2240
ACCTCCCCA/ACCCGAGGGACCCGACAGGCCGAGGGAAATAGAAGAAGAAGGGTGGAGAGAGACAGAGACA
2180 2190 2200 2210 2220 2230 2240
ACCTCCCCA/ACCCGAGGGACCCGACAGGCCGAGGGAAATAGAAGAAGAAGGGTGGAGAGAGACAGAGACA
8390 8400 8410 8420 8430 8440 8450
2260 2260 2270 2280 2290 2300 2310
GATGATTCTGATTAATGAAACGGATCCTTAGCACTTATCTGGGACGATCTGGGAGGCCTTGTGCGCTCTCAGC
2260 2260 2270 2280 2290 2300 2310
GATGATTCTGATTAATGAAACGGATCCTTAGCACTTATCTGGGACGATCTGGGAGGCCTTGTGCGCTCTCAGC

GATCCATTGAT TAGTGAACGATCCTGGCACTTATCTGGGACGATCTGCAGGCC-TGTGCCTCTTCAGC
 8460 8470 8480 8490 8500 8510 8520
 2320 2330 2340 2350 2360 2370 2380
 TACGCCCTTGTAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGAA
 :::::::::::::::::::::
 TACGCCCTTGTAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGAA
 8530 8540 8550 8560 8570 8580 8590
 2390 2400 2410 2420 2430 2440
 GCCCTCAAATATGGTGGAACTCCTACAGTATTGGAGTCAGGAACCTAAAG
 :::::::::::::::::::::
 GCCCTCAAATATGGTGGAACTCCTACAGTATTGGAGTCAGGAACCTAAAG
 8600 8610 8620 8630 8640 X

9. KUNZ-158-CL33. SCB

HIVSC Human immunodeficiency virus type 1, isolate SC (3

LOCUS HIVSC 4273 bp ss-RNA **VRL** 15-JUN-1988
DEFINITION Human immunodeficiency virus type 1, isolate SC (3' end of genome).
ACCESSION M17450
KEYWORDS
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate SC, provial DNA.
ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
REFERENCE 1 (bases 1 to 4273)
AUTHORS Gango, C., Guo, H.-G., Franchini, G., Aldovini, A., Collalti, E.,
 Farrell, K., Wong-Staal, F., Gallo, R. C. and Reitz, M. S. Jr.
TITLE Envelope sequences of two new United States HIV-1 isolates
JOURNAL Virology 164, 531-536 (1988)
STANDARD full staff_review
COMMENT Kindly made available in computer readable form by Marv Reitz,
 N. C. I., Bethesda, MD 20892 U.S.A. This isolate was taken from a
 California AIDS patient in 1984. There is an in-frame stop codon at
 position 3212 of the envelope coding sequence; the nef cds is
 uncertain beyond position 4049. A stop codon, 'taa,' in-frame with
 the nef sequence does exist at positions 4224-4226.

FEATURES	from	to/span	description
pept	< 1	330	vif protein (partial; AA at 1)
pept	270	560	vpr protein
pept	541	755	tat protein, exon 2 (first expressed exon)
	3088	3178	tat protein, exon 3 (AA at 3090)
pept	630	755	rev protein, exon 2 (first expressed exon)
	3089	3363	rev protein, exon 3 (AA at 3091)
pept	772	864	vpu protein (premature termination)
pept, ps	835	3505	envelope polyprotein (premature stop at 3212)
pept	3507	4226	nef protein
pre-msg	< 1 >	4273	genomic mRNA
pre-msg	< 1 >	4273	tat, rev, nef subgenomic mRNA
IVS	< 1	487	tat, rev, nef subgenomic mRNA intron 1
IVS	756	3088	tat cds intron 2
IVS	756	3088	rev cds intron 2
IVS	756	3088	tat, rev, nef subgenomic mRNA intron 2
LTR	3795 >	4273	3' LTR
rpt	4248 >	4273	R repeat 3' copy
cste	3212	3214	premature stop (tag) in env cds
BASE COUNT	1447 a	760 c	1053 g 1013 t

Initial Score = 1159 Optimized Score = 2139 Significance = 0.00
 Residue Identity = 87% Matches = 2194 Mismatches = 203
 Gaps = 97 Conservative Substitutions = 0

770 780 790 800 810 820 830
 TATGCCAAGTAGTATCAACTCAACTCTTTGAAATGGCAGTCAGCAGAAGAAGAGGTAGTAATTAGATCTGC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TATGCCAGTAGTATCAACTCACCTGCTGTTAAATGGCAGTCAGCAGAAGAAGAGGTAGTACTTAGGTCTGA
 1690 1700 1710 1720 1730 1740 1750
 840 850 860 870 880 890 900 910
 CAATTTCACAGAACATGCTAAACCCATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AAATTTCAACGACAAATGCTAAACCCATAATAGTACAGCTGAAAGAAGCTGTAGAAATTAATTGTACAAGGGCC
 1760 1770 1780 1790 1800 1810 1820 1830
 920 930 940 950 960 970 980
 CAACAAACAAATACAAAGAAAAGTATCGTATCCAGAGGGGACCCAGGGAGAGCATTGTTACAATAGGA---AA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CAACAAACAAATACAAAGAAAAGTATACATAT--AG---GACCAGGGAGAGCATTGTTACAACAGGAGACAT
 1840 1850 1860 1870 1880 1890
 990 1000 1010 1020 1030 1040 1050
 AATAGGAAATATGAAACAAAGC4CATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGATAGC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AATAGGAAATATGAAACAAAGC4CATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGATAGT
 1900 1910 1920 1930 1940 1950 1960
 1060 1070 1080 1090 1100 1110 1120
 TATCAAATTAAAGAGAACATTTGAAATAATAAAACAATACTTTAAGCAATCCTCAGGAGGGACCCAGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TATCAAATTAAAGAGAACATTTGAAATAATAAAACAATACTTTAATCGATCCTCAGGAGGGAGACCCAGA
 1970 1980 1990 2000 2010 2020 2030
 1130 1140 1150 1160 1170 1180 1190
 AATTGTAACGACAACTTTAATTGAGGGGAATTCTACTGTAATTCAACACAACTGTTAATAGTAC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AATTGTAACGACAACTTTAATTGAGGGGAATTCTACTGTAATTCAACACAACTGTTAGTAGTAC
 2040 2050 2060 2070 2080 2090 2100
 1200 1210 1220 1230 1240
 TTGTTTAATAGTACT--TGG---AGTA---CTGAAGG---GTCA-AATAACAC----TGAAG--GAAG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TTGTTTAATAGTACTGAAGGGTCAAATACACTGGAGGAAATGACACAAATCACCTCCCATGAGAATAAA
 2110 2120 2130 2140 2150 2160 2170
 1250 1260 1270 1280 1290 1300 1310
 TGAAACAAAT-CACA-CTCCGATCGAGAAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAAAGCAATG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AGAAATTAAACAAATCTGGAGGAGTAGGAAAGCAATGTATGCCCTCCCATCAA--AGGACAAG--TTA
 2180 2190 2200 2210 2220 2230 2240
 1320 1330 1340 1350 1360 1370 1380
 TATCCCCCTGGGATCGAGAAATAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAAGAGATGGT
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 ATTTTCAACAAAT--ATTACA---GGGCTGCT-AT-TAACAAAGAGATGG-TGGTAATAGCAAGA-ATGGT
 2250 2260 2270 2280 2290 2300
 1390 1400 1410 1420 1430 1440 1450 1460
 GGTAAATAACAAACAAATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AGCAAGAAATGAGAA---CACAGAGATCTTAGACCGGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAA
 2310 2320 2330 2340 2350 2360 2370
 1470 1480 1490 1500 1510 1520 1530
 TTATATAATATAAAGTAGTAAAGATGAAACCATTAGGAGTAGCACCCACCAAGGGCAAAGAGAAGAGTGTTG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TTATATAATATAAAGTAGTAAAGATGAAACCATTAGGAGTAGCACCCACCAAGGGCAAAGAGAAGAGAGTGTTG
 2380 2390 2400 2410 2420 2430 2440

1540 1560 1560 1570 1580 1590 1600
 CAGAGAGAAGAAAGAGACGTCGGGA---ATAGGAGCTTGTTCCTGGGTTCTGGGAGCAGCAGGAAGCACT
 1510 1530 1550 1570 1580 1590 1600
 CAGAGAGAAGAAAGAGACGTCGGGAACGATAGGAGCTATGTCCTGGGTTCTGGGAGCAGCAGGAAGCACT
 2450 2460 2470 2480 2490 2500 2510 2520
 1610 1620 1630 1640 1650 1660 1670
 ATGGAGCGACGCGTCAATGACCGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAAGCAGCAGAAC
 1620 1630 1640 1650 1660 1670 1680
 ATGGAGCGCTACGGTCAATGACCGCTGACGGTACAGGCCAGACTATTATTGTCTGGTATAGTGCAACAGCAGAAC
 2530 2540 2550 2560 2570 2580 2590
 1680 1690 1700 1710 1720 1730 1740
 AATTGTGCTTAAAGGGCTATTGAGGCCGCAACAGCACTGTGCAACTCACAGTCTGGGCACTCAAGCAGCTCCAG
 1690 1700 1710 1720 1730 1740 1750
 AATTGTGCTTAAAGGGCTATTGAGGCCGCAACAGCACTGTGCAACTCACAGTCTGGGCACTCAAGCAGCTCCAG
 2560 2570 2580 2590
 1750 1760 1770 1780 1790 1800 1810
 GCGAAATCCTGTCTGGAAAGATAACCTAAAGGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAAA
 1760 1770 1780 1790 1800 1810 1820
 GCGAAATCCTGTCTGGAAAGATAACCTAAAGGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAAA
 2570 2580 2590 2600 2610 2620 2630 2640 2650 2660
 1820 1830 1840 1850 1860 1870 1880 1890
 CTCAATTGGAAACACTCTGTGCGCTGGAAATGCTAGTTGGAGCTAATAAAATCTCTGGAACAGATTGGAAATAAC
 1830 1840 1850 1860 1870 1880 1890 1900
 CTCAATTGGAAACACTCTGTGCGCTGGAAATGCTAGTTGGAGCTAATAAAATCTCTGGAACAGATTGGGGTAAC
 2740 2750 2760 2770 2780 2790 2800
 1900 1910 1920 1930 1940 1950 1960
 ATBACCTGGATGCGTGGGAGAGAAATTACAAATTACACAGCTTAATACATTCCCTTAATTGAAGAATCG
 1910 1920 1930 1940 1950 1960 1970
 ATGAGCTTGTGCGTGGGAAAGAGAAATTGACAATTACACAGCTTAATATACACCTTAATTGAAGAATCG
 2810 2820 2830 2840 2850 2860 2870 2880
 1970 1980 1990 2000 2010 2020 2030
 CAAAACCAAGCAAGAAAAGAAATGAAACAGAAATTATTGGAATTAGATAAAATGGGCAAGTTGTGGAATTGGTTT
 1980 1990 2000 2010 2020 2030 2040
 CGGACCAACGAAAGAAATTGGAATTAGATAAAATGGGCAAGTTGTGGAATTGGTTT
 2890 2900 2910 2920 2930 2940 2950
 2040 2050 2060 2070 2080 2090 2100
 AACATAACAAATTTGGCTGTGCTATATAAAATATTCTATAATGATAGTAGGAGGGCTTGGTAGGTTAACAAATA
 2050 2060 2070 2080 2090 2100 2110
 AACATAACAAATTTGGCTGTGCTATATAAAATATTCTATAATGATAGTAGGAGGGCTTGGTAGGTTAACAAATA
 2060 2070 2080 2090 3000 3010 3020
 2110 2120 2130 2140 2150 2160 2170
 GTTITTTGGCTGTGCTATATAAGTGAATAGAGTTAGGCAGGGATATTCACTTACGTTTCAGACCCACCTC
 2120 2130 2140 2150 2160 2170 2180
 GTTITTTACGTGTACTTCTATAAGTGAATAGAGTTAGGCAGGGATATTCACTTACGTTTCAGACCCGCCTC
 3030 3040 3050 3060 3070 3080 3090
 2180 2190 2200 2210 2220 2230 2240 2250
 CCACCCCGAGGGAGCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGACAGAGACAGATCC
 2190 2200 2210 2220 2230 2240 2250 2260
 CCACCCCGAGGGAGCCGACAGGCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGACAGAGACAGATCC
 3100 3110 3120 3130 3140 3150 3160
 2260 2270 2280 2290 2300 2310 2320
 ATTGGATTAGTGAACGAGTCTTGTGCTTACGTTCTGGGACGATCTGCGGAGCCTTGTGCTCTTCAGCTACCA
 2270 2280 2290 2300 2310 2320 2330
 GGTGGATTAGTGAACGAGTCTTGTGCTTACGTTCTGGGACGATCTGCGGAGCCTTGTGCTCTTCAGCTACCA
 3170 3180 3190 3200 3210 3220 3230 3240

3330	3340	2350	2360	2370	2380	2390
CCCGTTGAAAGAATTACTCTT	GATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT					
CCCGTTGAAAGAATTACTCTT	GATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT					
3250	3260	3270	3280	3290	3300	3310
2400	2410	2420	2430	2440		
CAAAATTTGTTGAAATCTCTT	CAGTATTGAGTCAGGAACCTAAAG					
CAAAATTTGTTGAAATCTCTT	CAGTATTGAGTCAGGAACCTAAAG					
3320	3330	3340	3350	X		

10. KUNZ-158-HCL3S.3818

HIVMAL Human immunodeficiency virus type 1, isolate MAL,

LOCUS HIVMAL 3228 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate MAL, complete genome.
 ACCESSION M34166
 KEYWORDS .
 SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate MAL, proviral DNA, clone M-H11.
 ORGANISM Human immunodeficiency virus type 1
 Virusidae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
 REFERENCE 1 (bases 1 to 3228)
 AUTHORS Allain, M., Wain-Hobson, S., Montagnier, L. and Sonigo, P.
 TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients
 JOURNAL Cell 46: 63-74 (1986)
 STANDARD full staff_review
 COMMENT A region similar to the vpu cds of HIVELI exists from positions 1836 (starts with 'ata' instead of 'atg') to 5881. To date it has not been possible to construct an infectious clone for the MAL viral genome (Keith Pedersen, Johns Hopkins University School of Medicine, Baltimore, MD).
 FEATURES from to/span description
 pept 350 1867 gag polyprotein
 pept < 1863 4571 pol polyprotein (NH2-terminus uncertain; AA at 1863)
 pept 4616 5184 vif protein
 pept 5134 5424 vpr protein
 pept 5405 5613 tat protein, exon 2 (first expressed exon)
 pept 7969 8007 tat protein, exon 3 (AA at 7960)
 pept 5644 5819 rev protein, exon 2 (first expressed exon)
 pept 7959 8236 rev protein, exon 3 (AA at 7961)
 pept < 5636 5881 vpu protein (AA at 5636)
 pept 5789 6373 envelope polyprotein precursor
 pept 8400 8608 nef protein
 premmsg 1 8228 genomic mRNA
 premmsg 1 8228 tat, rev, nef subgenomic mRNA
 IVS 306 5351 tat, rev, nef subgenomic mRNA intron 1
 IVS 5620 7558 tat cds intron 2
 IVS 5620 7958 rev cds intron 2
 IVS 5620 7558 tat, rev, nef subgenomic mRNA intron 2
 LTR < 1 !77 5' LTR
 LTR 8675 > 8228 3' LTR
 rpt < 1 95 R repeat 5' copy
 rpt 8133 8228 R repeat 3' copy
 binding 179 196 primer (Lys-tRNA) binding site
 signal 8204 8203 mRNA polyadenylation signal
 BASE COUNT 3356 a 1627 c 2204 g 2043 t
 ORIGIN Dep. site of genomic RNA.

Initial Score = 916 Optimized Score = 2041 Significance = 0.00
 Residue Identity = 80% Matches = 2066 Mismatches = 349

X 10 20 30 40 50 60 70
ATTCAGACTTAAAGGAGAAATACTACGCACTTGTTGGAGATGGGGGTGGAAATGGGCACCATGCTCCTGGGATA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATTCAGACTTAAAGGAGAAATACTACGCACTTGTTGGAGATGGGGGTGGAAATGGGCACCATGCTCCTGGGATA
5800 5810 5820 5830 5840 5850 5860
80 160 100 110 120 130 140
TTCTATGATCTATAGTCTACAGAAGAAATTGTCGGTCACTGCTATTATGGGGTACCTGTGTGGAAAGGAAGCA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TTCTATGACUTGTAGTATTGCAAGAAGATTGTGGGTTACAGTTATTATGGGGTACCTGTGTGGAAAGGAAGCA
5870 5880 5890 5900 5910 5920 5930
150 160 170 180 190 200 210
ACGCCCCCTCTTTTGTGATCAGATECTAAAGCCATATGATAACAGAGGTACATAATGTTTGGGCCACACAT
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ACGACTACTCTATTTTGTGATCAGATECTAAATGATAATGAAACAGAAGTACATAACATCTGGGCTACACAT
5940 5950 5960 5970 5980 5990 6000 6010
220 230 240 250 260 270 280
GCCTGTGTAACCCACACGACCGAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGG
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GCCTGTGTAACCCACACGACCGAACCCACAAGAAGTAGACTGGAAAATGTGACAGAAGGGTTAACATGTGG
6020 6030 6040 6050 6060 6070 6080
290 300 310 320 330 340 350 360
AAAATGAACTGATGAAACATGATGAGGATATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAAATGAACTGATGAAACATGATGAGGATATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
6090 6100 6110 6120 6130 6140 6150
370 380 390 400 410 420
AAATTAACCCCCTGATCTTAAAGTGCACGTGATTG---GGGAATGCTACTAAT---ACCAACT
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAGCTAACCCCCTGATCTTAAAGTGCACGTGAACTATGTGAATGGGACTGCTGTGAATGGGACTAATGCT
6160 6170 6180 6190 6200 6210 6220
430 440 450 460 470 480 490
AGTAATACCAATACTGATGAGGAAATGTGATGAGGAGAGATAAAAAACTGCTCTTCAATATC
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GGGAGT---ATAAGGACTAATGCAAGATTGGAAAATGGGAGAAGTGAAAAGTGCCTCTTCAATATA
6230 6240 6250 6260 6270 6280 6290
500 510 520 530 540 550 560
ACGCAAGNATAAGGCTAAATGCTGAGAAAGATATGCATTTTTATAAACCTTGATATAATACCAATA---
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ACCCAGTAAAGGAGTGTATAAAAGGC---AAGAATATGCAACTTTATAACCTTGATCTAGTACAAATAGAT
6300 6310 6320 6330 6340 6350 6360
570 580 590 600 610 620 630
GATAATGATACTACCGCTATACTGACAAGTTGACACCTCAGTCATTACACAGGGCTGTCCAAAGGTA
||||| ||||| ||||| ||||| ||||| ||||| |||||
GATAATGATAATAGTGTATAAGGCTAATAAAATTGTAATACCTCAGTAATTACACAGGGCTGTCCAAAGGTA
6370 6380 6390 6400 6410 6420 6430
640 650 660 670 680 690 700 710
TCCTTTGAGCCAAATTCGGATACATTATTGTCGGCCCCGGCTGGTTGCGATTCTAAATGTAATAATAAGACG
||||| ||||| ||||| ||||| ||||| ||||| |||||
ACCTTTGATGCAATTCCGATACATGATTGTGGCCCGAGCTGGTTGCAATTCTAAAGTGTAAATGATAAGAAG
6440 6450 6460 6470 6480 6490 6500
720 730 740 750 760 770 780
TTCAATGGAAACATGACCGATGACGAACTGCAAGTACAATGTACACATGGAATTAGGCCAGTAGTATCA
||||| ||||| ||||| ||||| ||||| ||||| |||||
TTCAATGGAAACATGACCGATGACGAACTGCAAGTACAATGTACACATGGAATTAGGCCAGTAGTATCA

6510 6520 6530 6540 6550 6560 6570 6580
780 800 810 820 830 840 850
ACTCAACTGCTGTTGAAATGGCAGTCAGCAGAAGAGAGGGTAGTAATTAGATCTGCCAATTCACAGACAAT
|||||:|||||:|||||:|||||:|||||:|||:|||||:|||:|||||:
ACTCAACTGCTGTTAAATGGCAGTCAGCAGAAGAGAGATAATGATTAGATCTGAAAATCTCACAGACAAT
6590 6600 6610 6620 6630 6640 6650

860 870 880 890 900 910 920
GCTAAAACCATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAAACAACAATACAAGA
|||||:|||||:|||:|||:|||||:|||||:|||:|||||:
ACTAAAAAACATAATAGTACAGCTTAATGAAACTGTACAACATTAAATTGTACAAGGCCTGGAAACAATACAAGA
6660 6670 6680 6690 6700 6710 6720

930 940 950 960 970 980 990
AAAAGTATCCGTAATCCAGAGGGACCGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAATATGAGACAA
|||||:|||:|||:|||:|||:|||:|||:|||:|||:
AGAQGGATACATTTG-----GCCCGAGGGCAAGCACTCTACAAACAGGGATAGTAGGAGATATAAGAAGA
6730 6740 6750 6760 6770 6780 6790

1000 1010 1020 1030 1040 1050 1060 1070
GCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAATTAAGAGAACAA
|||||:|||:|||:|||:|||:|||:|||:
GCATATTGTACTATTAATGAAACAGAATGGGATAAAACTTTACAAACAGGTAGCTGTAACACTAGGA-AGC--
6800 6810 6820 6830 6840 6850 6860

1080 1090 1100 1110 1120 1130 1140
TTTGGAAAATAATAAAACAATAATCTTAAAGCAATCCTCAGGAGGGGGACCCAGAAATTGTAACGCACAGTTTT
|||:|||:|||:|||:|||:
CTTCTTAACAAACAAACAAATAATTAAATTCACTCCTCAGGAGGGGGACCCAGAAATTACAACACACAGTTTT
6870 6880 6890 6900 6910 6920 6930

1150 1160 1170 1180 1190 1200 1210
AATTGTTGGAGGGGAATTTTCTACTGTAATTCAACACAACTGTTAATAGTACTTGGTTAATAGTACTTGG
|||||:|||||:|||||:
AATTGTTAGAGGGGAATTTTCTACTGTAATACATCAAAACTGTTAATAGTACATGGCAGAATAAT-GGTGC
6940 6950 6960 6970 6980 6990 7000

1220 1230 1240 1250 1260 1270 1280
AGTACTGAAGGGCTAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTATA
|||||:|||:
AAGACT-AAG---TAATAGCACAGAGTCAACTGGTAGTACACTCCCATGCAGAATAAAACAATTATA
7010 7020 7030 7040 7050 7060 7070

1290 1300 1310 1320 1330 1340 1350
AACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCATCAAAT
|||:
AATATGTGGCAGAAAACAGGAAAAGCTATGTATGCCCTCCCATCGCAGGAGTCATCAACTGTTATCAAAT
7080 7090 7100 7110 7120 7130 7140

1360 1370 1380 1390 1400 1410 1420
ATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACA---ACAAT-GGGTC---CGAGATCTTCAGACCT
|||||:
ATTACAGGGCTGCTATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATGAGACCTTAAGACCT
7150 7160 7170 7180 7190 7200 7210

1430 1440 1450 1460 1470 1480 1490
GGAGGAGGGAGATATGAGGGACAAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAAATTGAACCCATTA
|||||:
GGAGGAGGGAGATATGAGGGACAAATTGGATAAGTGAATTATATAAATATAAAGTAGTAAAGAATTGAACCCCTA
7220 7230 7240 7250 7260 7270 7280

1500 1510 1520 1530 1540 1550 1560
GGAGTAGCACCCACCAAGGCAAAAGAGAAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTG
|||||:
GGAGTAGCACCCACCAAGGCAAAAGAGAAGTAGTGGAGAAGAGAAAAAGAGCAATAGGACTAGGAGGCCATG

7290	7300	7310	7320	7330	7340	7350	
1570	1580	1590	1600	1610	1620	1630	1640
TTCCCTTGGGTTC	TTGCAAGCAGCAUGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCAGA						
TTCCCTTGGGTTC	TTGCAAGCAGCAUGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCAGA						
7360	7370	7380	7390	7400	7410	7420	7430
1650	1660	1670	1680	1690	1700	1710	
CAATTATTGTC166	TATAGTGCAACAGCAGAACAAATTGCTGAGGGCTATTGAGGCAGAACAGCATCTGTTG						
CAGTTACTGTC	TGGTATAGTGCAACAGCAAAACAATTGCTGAGGGCTATAGAGGCAGAACAGCATCTGTTG						
7440	7450	7460	7470	7480	7490	7500	
1720	1730	1740	1750	1760	1770	1780	
CAACTCACACGTC	TGGGTCATCAECAGCTCCAGGCAAGAACATCCTGGCTGTGGAAAGATAACCTAACAGGATCAA						
CAACTCACACGTC	TGGGTCATCAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGATAACCTAACAGGATCAA						
7510	7520	7530	7540	7550	7560	7570	
1790	1800	1810	1820	1830	1840	1850	
CAGCTCCTGGGATT	TGGGTTGCTCTGGAAAAGTCATTGCAACCACGTGCTGTGCCTTGGAAATGCTAGTTGG						
CAGCTCCTAGGAAT	TGGGTTGCTCTGGAAAACACATTTGCAACACATTGCTGACACATTGTGCCTTGGAACTCTAGTTGG						
7580	7590	7600	7610	7620	7630	7640	
1860	1870	1880	1890	1900	1910	1920	
AGTAATAAATCTG	TAACAGATTGCAATAACATGACCTGGATGGAGTGGGACAGAGAAAATTAAACAATTAC						
AGTAATAGATCTG	TAACATGACATTGGAAATAATGACCTGGATGCAGTGGGAAAAAGAAATTAGCAATTAC						
7650	7660	7670	7680	7690	7700	7710	
1930	1940	1950	1960	1970	1980	1990	2000
ACAAAGCTTATAUAT	TGCTTAATTGAAAGAACATCGCAAAACAGCAAGAAAAGAACATTGAAACAAAGAACATTATTGGAA						
ACAAAGCTTATAUAT	TGCTTAATTGAAAGAACATCGCAAAATCCAGCAAGAAAAGAACATTGAAACAAAGAACATTATTGGAA						
7720	7730	7740	7750	7760	7770	7780	7790
2010	2020	2030	2040	2050	2060	2070	
TTACATAAATG	GGGTTGTGCAATTGGTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATA						
TTGCAACAAATG	GGGTTGTGCAATTGGTTAACATAACAAATTGGCTGTGGTATATAAAATATTCTATA						
7800	7810	7820	7830	7840	7850	7860	
2080	2090	2100	2110	2120	2130	2140	
ATGAATAGTGGACCT	TTGCAAGGAAATTGGCTGTGACTTTCTATAGTGAATAGAGTTAGGCAG						
ATATGAGTGGACCT	TTGCAAGGAAATTGGCTGTGACTTTCTATAGTGAATAGAGTTAGGCAG						
7870	7880	7890	7900	7910	7920	7930	
2150	2160	2170	2180	2190	2200	2210	
GGAAATTGCAACATTA	UATTGCAACCCACCTCCCAACCCCCGAGGGG---ACCCGACAGGCCGAAGGAATA						
GGATACTGACCTGCTG	CAACCCCTCCCAACACCGAGGGGACCCGACAGGCCGAAGGAATA						
7940	7950	7960	7970	7980	7990	8000	
2220	2230	2240	2250	2260	2270	2280	
GAAGNAGAATG	GGGAAATTGCAACCCCTCCCAACACCGAGGGGACCCGACAGGCCGAAGGAATA						
GAAGNAGAATG	GGGAAATTGCAACCCCTCCCAACACCGAGGGGACCCGACAGGCCGAAGGAATA						
8010	8020	8030	8040	8050	8060	8070	
2290	2300	2310	2320	2330	2340	2350	
GACCATGTC	TTGCAACCCCTCCCAACACCGAGGGGACCCGACAGGCCGAAGGAATA						
GACCATGTC	TTGCAACCCCTCCCAACACCGAGGGGACCCGACAGGCCGAAGGAATA						

8080 8080 8100 8110 8120 8130 8140
2360 2370 2380 2390 2400 2410 2420 2430
TGTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCA
TGTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAAATCTCCTGCAATATTGGGGTCA
8150 8160 8170 8180 8190 8200 8210 8220

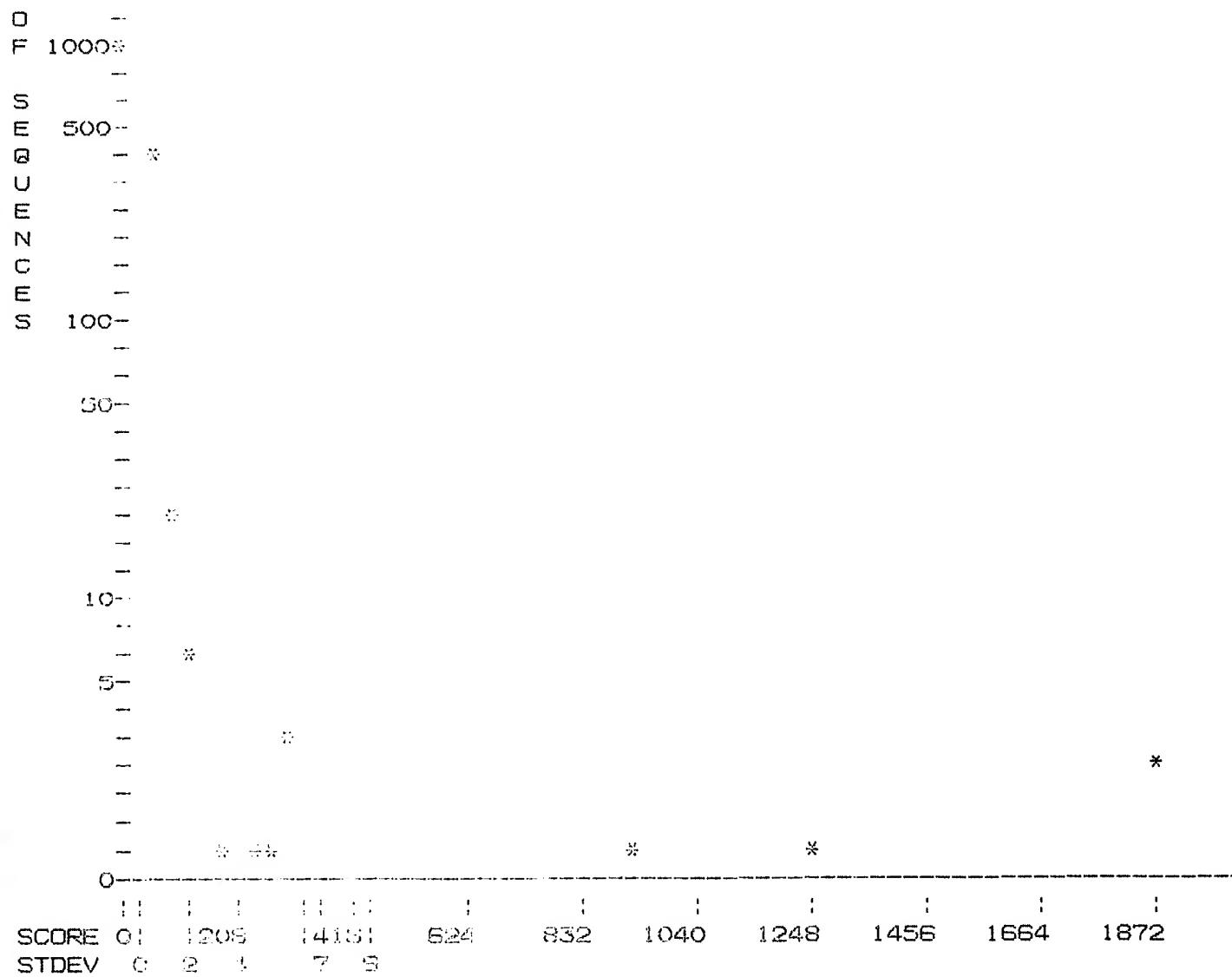
X
GAAACTAAAAAG
GAAACTGAAAG
8230

Results file kunz-158-cl33umb1.res made by sheppard on Thu 8 Mar 90 11:32:28-PST.

Query sequence being compared: KUNZ-158-CL33. SEQ
Number of sequences searched: 3460
Number of scores above cutoff: 10

Results of the initial comparison of KUNZ-158-CL33. SEQ with:
Data Bank : UEMBL 21, all entries

10000-
-
N -
U 5000-
M -
B -
E -->
R -



PARAMETERS

Similarity metric	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	145		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation		
	40	38	46.71		
Times:	CPU		Total Elapsed		
	00:07:26.04	00:32:23.00			
Number of residues:	5125898				
Number of sequences searched:	3460				
Number of scores above cutoff:	10				

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.				
		Length	Score	Score	Sig.	Frame
**** 35 standard deviations above mean ****						
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2176	39.22	0
2. HIVH309	Human T-cell lymphotropic virus	9749	1872	2176	39.22	0
**** 25 standard deviations above mean ****						
3. HIVELICS	Human lymphadenopathy virus (E)	9176	1246	1893	25.82	0
**** 18 standard deviations above mean ****						
4. HIVMAL08	Human lymphadenopathy virus (M)	9229	916	2041	18.75	0
**** 5 standard deviations above mean ****						
5. HIV2R00X	Human immunodeficiency virus t	9671	306	1185	5.69	0
6. REHTLV4C	HTLV-3 (HTLV-4) partial provir	5391	299	1247	5.54	0
7. RESIVAXX	Simian immunodeficiency virus	9264	294	1240	5.44	0
8. REHTVM00X	Simian immunodeficiency virus	9646	290	1240	5.35	0
**** 4 standard deviations above mean ****						
9. RESIV261	Simian Immunodeficiency virus	1142	259	565	4.69	0
**** 3 standard deviations above mean ****						
10. M15127	Figure 1. Structure of the art	306	184	298	3.08	0

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.				
		Length	Score	Score	Sig.	Frame
**** 35 standard deviations above mean ****						
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2176	0.00	0
2. HIVH309	Human T-cell lymphotropic virus	9749	1872	2176	0.00	0
3. HIVMAL08	Human lymphadenopathy virus (M)	9229	916	2041	0.00	0
4. HIVELICS	Human lymphadenopathy virus (E)	9176	1246	1893	0.00	0
5. REHTLV4C	HTLV-3 (HTLV-4) partial provir	5391	299	1247	0.00	0
6. REHTVM00X	Simian immunodeficiency virus	9646	290	1240	0.00	0
7. RESIVAXX	Simian immunodeficiency virus	9264	294	1240	0.00	0
8. HIV2R00X	Human immunodeficiency virus t	9671	306	1185	0.00	0
9. RESIV261	Simian Immunodeficiency virus	1142	259	565	0.00	0
10. M15127	Figure 1. Structure of the art	306	184	298	0.00	0

SEARCH STATISTICS

Scores: Mean: 2176 Median: 2177 Standard Deviation: 0.00

Times: CPU: 00:00:00.00 Total Elapsed: 00:00:00.00

Number of residues: 73322

Number of sequences optimized: 10

1. KUNZ-196-CL43. SEQ

REHTLV3 Human T-cell leukaemia type III (HTLV-III) provira

ID REHTLV3 standard t RNA: 9748 BP.

XX

AC XC1782

XX

DT 05-SEP-1987 (am correction)
 DT 01-SEP-1987 (am correction)
 DT 03-AUG-1987 (am correction)
 DT 24-OCT-1988 (minor modification)
 DT 05-NOV-1988 (KW added)
 DT 28-MAR-1989 (first entry)
 XX
 DE Human T-cell leukaemia type III (HTLV-III) proviral genome
 DE (AIDS virus for acquired immune deficiency syndrome)
 XX
 KW acquired immune deficiency syndrome; direct repeat; endonuclease;
 KW glycoprotein; inverted repeat; protease; provirus;
 KW reverse transcriptase; terminal repeat.
 XX
 OS Human T-cell leukaemia virus type III
 OC Viridae; ds-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-3748)
 RA Ratner L., Haseltine W., Paterca R., Livak K. J., Starcich B. R.,
 RA Josephs S. F., Noran E. R., Rafalski J. A., Whitehorn E. A.,
 RA Baumleiter K., Ivanoff L., Petteway S. R. Jr., Pearson M. L.,
 RA Leutenberger J. A., Papas T. S., Ghayeb J., Chang N. T., Gallo R. C.,
 RA Wong-staal F.;
 RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
 RL Nature 313:277-284(1985).
 XX
 RN [2]
 RA Muesing M. A., Smith D. H., Cabradilla C. D. JR., Benton C. V.,
 RA Kaszy L. A., Capon D. J.;
 RT "Nucleic acid structure and expression of the human AIDS/
 RT lymphadenopathy retrovirus";
 RL Nature 313:450-458(1985).
 XX

FH	Key	From	To	Description
FT	INVREP	1	2	inverted repeat
FT	SITE	1	634	long terminal repeat
FT	PRM	427	430	TATA-box
FT	SITE	453	453	U3 region
FT	CAP	454	454	cap site
FT	SITE	454	551	R region
FT	SITE	552	634	U5 region
FT	INVREP	633	634	inverted repeat
FT	SITE	635	653	tRNA binding site (tRNA-Lys)
FT	CDS	787	1162	gag p17
FT	CDS	787	2321	gag precursor polypeptide
FT	CDS	1183	2321	gag p24 and gag p15 for major capsid protein and for put. retroviral nucleic acid binding protein (NBP)(ref. 2) (boundaries not defined)
FT				
FT	RPT	1588	2007	direct repeat
FT	RPT	2031	2065	direct repeat
FT	CDS	2001	5125	pol precursor polypeptides put. protease at 5' terminus reverse transcriptase
FT				
FT	RPT	2123	2163	put. endonuclease at 3' terminus
FT	RPT	2164	2176	direct repeat
FT	CDS	5040	5648	SDR short open reading frame put. vestigial env gene
FT	CDS	6373	8821	env-lor precursor polypeptide
FT	CDS	6323	8821	envelope glycoprotein
FT	SITE	7738	7787	put. peptide cleavage site
FT	CDS	7767	8821	put. lor transmembrane protein

FT SITE 8093 8103 poly purine stretch
 FT SITE 9116 9567 U3 region
 FT RPT 9116 9740 long terminal repeat
 FT SITE 9568 9695 R region
 FT SITE 9641 9646 polyadenylation signal
 FT SITE 9660 9743 U3 region
 FT INVREP 9747 9749 inverted repeat
 XX
 SQ Sequence 9748 BP; 3431 A; 1781 C; 2368 G; 2168 T; 0 other;
 Initial Score = 1872 Optimized Score = 2176 Significance = 0.00
 Residue Identity = 85% Matches = 2243 Mismatches = 150
 Gaps = 106 Conservative Substitutions = 0

X	10	20	30	40	50	60	
ATGAGAGTGA-AGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGAAATGGGGCAC-CATGCTCCTTGG							
CTAATAGAAAGAGCAAAAGACATGGCAAT----GAGAGTGAGGAGAAATATCAGCACTTGTGGAGATGGG	6230	6240	6250	6260	6270	6280	6290
70	80	90	100	110	120		
GATAATTCATGCTACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--							
GGTAGAGATGGCAACCCATECTGTTGGATGTTGATCTGTAGTGCTACAGAAAATTGTGGGTACAG	6300	6310	6320	6330	6340	6350	6360
130	140	150	160	170	180		
-CT---GTGTGCAA-----GGAAGCAA-CCACCA-CTCTATTTGTGCATCAGATGCTAAAGCATATGAT							
TCTATTATGGGATACCTGTGTCAGAAGGAAGCAACCACACTCTATTTG-----TGC---ATCAGATGCT	6370	6380	6390	6400	6410		6420
190	200	210	220	230	240		
A---CAGAGG-TACATA---A1---GTTTGGGCCACACATGCCTG--TGTACCCACAGA-CCCCAACCCAC							
AAACGATATGATACAGAGGTACATAATGTTGGC-CACA--CATGCCTGTGTAACCCACAGACCCCAACC-C	6430	6440	6450	6460	6470	6480	6490
250	260	270	280	290	300	310	
AAGAAGTAACTATTGCTAAATGTGACAGAGAAAATTAAACATGTGGAAA--ATGACATGGTAGAACAGATG-C							
A-CAGAACGTA---GTA-TTG---GTAAAT-GTGAACA---GAAAATTAAACAT-GTGGAA-AAATGAC	6500	6510		6520	6530	6540	6550
320	330	340	350	360	370	380	
ATGAGGATATATCAG-TTTATG-GCACTAAAGCTAAAGCCATGTG-TAAAAA--TTAACCCCACCTCTGTGT							
AT---GGTAGAA-CAGATGUATGAGGATATAA---TCAGTTTATGGGATCAAAGCCTAAAGCCA---TGTGT	6560	6570		6580	6590	6600	6610
390	400	410	420	430	440		
TAGTTTAA-AGTGCACTGAATTGG---GGAATGCTACTAAT---ACCAACTAGTAATACCAATAGTAGTA							
AAAATTAAACCCACTGTGTGTTAGTTAAAGTGC-ACTGATTGAAAGAATGATACTAATACCAATAGTAGTA	6620	6630	6640	6650	6660	6670	6680
450	460	470	480	490	500	510	
GCGGAGAAATGATGATGGAGAAAGGAGAGATAAAAACGTCTTCAATATCAGCACAAGNATAAGAGGTA							
GCGGAGAAATGATAATGGAGAAAGGAGAGATAAAAACGTCTTCAATATCAGCACAAGCATAAGAGGTA	6650	6700	6710	6720	6730	6740	6750
520	530	540	550	560	570	580	
AGGTGCAAAAGAAATAAGCATTTTATGAAACCTTCATATACCAATAGATAATGATACTACCACTATA							
AGGTGCAAAAGAAATAAGCATTTTATGAAACCTTCATATACCAATAGATAATGATACTACCACTATA							

6720 6770 6780 6790 6800 6810 6820
590 600 610 620 630 640 650 660
CGTTCACAAAGTTCTAACACCTCACTCATTACACAGGGCTGTCCAAAGGTATCCTTGAGCCAATTCCCATA
CGTTCACAAAGTTCTAACACCTCACTCATTACACAGGGCTGTCCAAAGGTATCCTTGAGCCAATTCCCATA
6030 6040 6050 6060 6070 6080 6090 6100
670 680 690 700 710 720 730
ATTATTGTCGGGGGGCGGGTTTTGGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
ATTATTGTCGGGGGGCGGGTTTTGGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
6900 6910 6920 6930 6940 6950 6960 6970
740 750 760 770 780 790 800
CAAATGTCAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
CAAATGTCAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA
6980 6990 7000 7010 7020 7030 7040
810 820 830 840 850 860 870
GTCTAECAAATCTGTAGAAATTAAATTAGATCTGCCAATTTCACABACAATGCTAAAACCATAATAGTACAGC
GTCTAGGCAGAAAGAGGGTAGTAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAATAGTACAGC
7050 7060 7070 7080 7090 7100 7110
880 890 900 910 920 930 940
TGAACCAATCTGTAGAAATTAAATTAGTACAAAGACCCAAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
TGAACCAATCTGTAGAAATTAAATTAGTACAAAGACCCAAACAACAATACAAGAAAAAGTATCCGTATCCAGAGAG
7120 7130 7140 7150 7160 7170 7180
950 960 970 980 990 1000 1010 1020
GACCGAGGGAGGCAATTGGTACAAATTAGGAAAAATAGGAAATATGAGACAAAGCACATTGTAACATTAGTAGAG
GACCGAGGGAGGCAATTGGTACAAATTAGGAAAAATAGGAAATATGAGACAAAGCACATTGTAACATTAGTAGAG
7190 7200 7210 7220 7230 7240 7250
1030 1040 1050 1060 1070 1080 1090
CAAAATGCAATCTCACATTAAACAGATAGCTAGCAAATTAAAGAGAAACAATTGGAAATAATAAAACAATAA
CAAAATGCAATCTCACATTAAACAGATAGCTAGCAAATTAAAGAGAAACAATTGGAAATAATAAAACAATAA
7260 7270 7280 7290 7300 7310 7320 7330
1100 1110 1120 1130 1140 1150 1160
TCTTTAACGCAATTCTCAGGGGGGACCCACAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
TCTTTAACGCAATTCTCAGGGGGGACCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
7340 7350 7360 7370 7380 7390 7400
1170 1180 1190 1200 1210 1220 1230
ACTGTAATTCAACACAACTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
ACTGTAATTCAACACAACTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAAAGGGTCAAATAACA
7410 7420 7430 7440 7450 7460 7470
1240 1250 1260 1270 1280 1290 1300
CTGAAGGGAGGTCAACACAACTCACACTCCATGCAGAAATAAACAAATTATAAACATGTGGCAGGAAGTAGGAA
CTGAAGGGAGGTCAACACAACTCACACTCCATGCAGAAATAAACAAATTATAAACATGTGGCAGGAAGTAGGAA
7480 7490 7500 7510 7520 7530 7540
1310 1320 1330 1340 1350 1360 1370 1380
AAGCAATGTAATTGGGGTCCCATCAAGGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA
AAGCAATGTAATTGGGGTCCCATCAAGGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAAACAA

8340	8350	8360	8370	8380	8390	8400	8410
2180	2190	2200	2210	2220	2230	2240	
ACCTCCCCAACCCGAAAGGACCCGACAGGCCGAAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA							
ACCTCCCCAATCGGGAAGGACCCGACAGGCCGAAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA							
8420	8430	8440	8450	8460	8470	8480	
2250	2260	2270	2280	2290	2300	2310	
GATCUATTGATTAATGAAACGGATCCTTACCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGC							
GATCUATTGATTAATGAAACGGATCCTTACCACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGC							
8490	8500	8510	8520	8530	8540	8550	
2320	2330	2340	2350	2360	2370	2380	
TACCAACCGCTTCAAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAA							
TACCAACCGCTTCAAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAA							
8560	8570	8580	8590	8600	8610	8620	
2390	2400	2410	2420	2430	2440		
GCCCTCAAATATTGGAAATCTCCTACAGTATTGGAGTCAGGAACCTAAAG							
GCCCTCAAATATTGGTGAATCTCCTACAGTATTGGAGTCAGGAGCTAAAG							
8630	8640	8650	8660	8670	X		

2. KUNZ-158-CL33. SEQ

HTVH3CG Human T-cell lymphotropic virus type III, complete

ID HTVH3CG standard; RNA; 9749 BP.
 XX
 AC KC20101; K02008; K02009;
 XX
 DT 09-SEP-1987 (am correction)
 DT 03-SEP-1987 (am correction)
 DT 03-SEP-1987 (am correction)
 DT 02-SEP-1987 (am correction)
 DT 01-SEP-1987 (am correction)
 DT 23-JUN-1987 (minor modifications)
 DT 28-OCT-1989 (incorporated)
 XX
 DE Human T-cell lymphotropic virus type III, complete reference genome
 DE (isolates HXB2, HXB3, BH10, BHS and BH8 of HTLV-III DNA).
 XX
 KW acquired immune deficiency syndrome; complete genome; env gene;
 KW gag gene; long terminal repeat; pol gene; polyprotein; provirus;
 KW reverse transcriptase; TAR protein; trans-activator.
 XX
 OS Human T-cell lymphotropic virus type III
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-853, 8116-8743)
 RA Starcich B. R., Rother L., Josephs S. F., Okamoto T., Gallo R. C.,
 RA Wong-staal F.;
 RT "Characterization of long terminal repeat sequences of htlv-iii";
 RL Science 227:538-540(1985).
 XX
 RN [2] (bases 1-8749)
 RA Rother L., Haseltine W., Patarca R., Livak K. J., Starcich B. R.,
 RA Josephs S. F., Doreen E. R., Rafalski J. A., Whitehorn E. A.,
 RA Baumlester K., Ivanoff L., Petteway S. R. Jr., Pearson M. L.,
 RA Lautenberger J. A., Papes T. S., Ghrayeb J., Chang N. T., Gallo R. C.,
 RA Wong-staal F.;
 RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
 RL Nature 317:277-284(1985).

- XX
- RN [3] (bases 508-5666) exons only, tat mRNA
RA Arya S. K., Guo C., Josephs S. F., Wong-staal F.;
RT "Trans-activator gene of human T-lymphotropic virus type III
(HTLV-III)";
RL Science 233:65-73(1985).
- XX
- RN [4] (bases 5775-6082, 8387-8499)
RA Sodroski J. G., Paterca R., Rosen C. A., Wong-staal F., Haseltine W.;
RT "Location of the trans-activating region on the genome of human
T-cell lymphotropic virus type III";
RL Science 233:74-77(1985).
- XX
- RN [5] mRNA splice sites
RA Rabson A. B., Daugherty D. F., Venkatesan S., Boulukos K. e.,
RA Benn S. J., Folks T. M., Feorino P., Martin M.;
RT "Transcription of novel open reading frames of AIDS retrovirus
during infection of lymphocytes";
RL Science 233:1388-1390(1985).
- XX
- RN [6] 27k antigen cds
RA Allan J. S., Calligan J. E., Lee T. H., McLane M. F., Kanki P. J.,
RA Groopman J. E., Essex M.;
RT "A new HTLV-III/LAV encoded antigen detected by antibodies from
AIDS patients";
RL Science 230:810-813(1985).
- XX
- RN [7] (bases 5778-8933) in hxb-3
RA Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R.,
RA Shaw G., Wong-staal F., Reddy E. P.;
RT "HTLV-III env gene products synthesized in E. coli are recognized
by antibodies present in the sera of AIDS patients";
RL Cell 41:879-886(1985).
- XX
- RN [8] gp160 and gp120 coding sequences
RA Allan J. S., Calligan J. E., Barin F., McLane M. F., Sodroski J. G.,
RA Rosen C. A., Haseltine W. A., Lee T. H., Essex M.;
RT "Major glycoprotein antigens that induce antibodies in AIDS
patients are encoded by HTLV-III";
RL Science 233:1091-1094(1985).
- XX
- RN [9] regulatory sequences in the itr
RA Rosen C. A., Sodroski J. G., Haseltine W. A.;
RT "The location of cis-acting regulatory sequences in the human T
cell lymphotropic virus type III (HTLV-III/LAV) long terminal
repeat";
RL Cell 41:813-823(1985).
- XX
- RN [10] (bases 1-9749)
RA Van Beveren C., Van Beveren C., Van Beveren C.;
RT "Appendix B: HTLV-3/LAV genome";
RL (in) Weiss R., Teich N., Varmus and Coffin J. M. (eds.);
RL RNA Tumor Viruses Second Edition:1102-1148
RL Cold Spring Harbor Laboratory, New York (1985)
- XX
- RN [11] trans-activator function and tar sequence
RA Rosen C. A., Sodroski J. G., Goh W. C., Dayton A. I., Lippke J.,
RA Haseltine W. A.;
RT "Post-transcriptional regulation accounts for the trans-activation
of the human T-lymphotropic virus type III";
RL Nature 319:555-559(1986).
- XX
- RN [12] pol coding sequence
RA Marzo Veronese F., Copeland T. D., DeVico A. L., Rahman R.,
RA Ovineszian S., Gallo R. C., Sarnegadharan M. G.;
RT "Characterization of highly immunogenic p66/p51 as the reverse

- RT transcriptase of HTLV-III/LAV";
RL Science 231:1288-1291(1986).
- XX
- RN [13] the 23k sor gene product
RA Kan N. C., Franchini G., Wong-staal F., DuBois G. C., Robey W. G.,
RA Lautenberger J. A., Papas T. S.;
RT "Identification of HTLV-III/LAV sor gene product and detection of
RT antibodies in human sera";
RL Science 231:1553-1555(1986).
- XX
- RN [14] pol nh2-terminal region
RA Kramer R. A., Scheber M. D., Skalka A. M., Ganguly K., Wong-staal F.,
RA Reddy E. P.;
RT "HTLV-III gag protein is processed in yeast cells by the virus
RT pol-protease";
RL Science 231:1580-1584(1986).
- XX
- RN [15] sor 23K protein
RA Lee T. H., Colligan J. E., Allan J. S., McLane M. F., Groopman J. E.,
RA Essex M.;
RT "A new HTLV-III/LAV protein encoded by a gene found in cytopathic
RT retroviruses";
RL Science 231:1546-1549(1986).
- XX
- RN [16] sor 23k protein
RA Sodroski J. G., Goh W. C., Rosen C. A., Tartar A., Portetelle D.,
RA Burny A., Haseltine W.;
RT "Replicative and cytopathic potential of HTLV-III/LAV with sor
RT gene deletions";
RL Science 231:1549-1553(1986).
- XX
- RN [17] sp1 binding sites in the promoter region
RA Jones K. A., Kadonaga J. T., Luciw P. A., Tjian R.;
RT "Activation of the AIDS retrovirus promoter by the cellular
RT transcription factor, Sp1";
RL Science 232:755-759(1986).
- XX
- RN [18] acceptor and donor splice sites for tat and 27k
RA Arya S. K., Gallo R. C.;
RT "Three novel genes of human T-lymphotropic virus type III: Immune
RT reactivity of their products with sera from acquired immune
RT deficiency syndrome patients";
RL Proc. Natl. Acad. Sci. U. S. A. 83:2209-2213(1986).
- XX
- RN [19] deletion mutants in the tat gene
RA Dayton A. I., Sodroski J. G., Rosen C. A., Goh W. C., Haseltine W. A.;
RT "The trans-activator gene of the human T cell lymphotropic virus
RT type III is required for replication";
RL Cell 44:841-847(1986).
- XX
- RN [20] hypervariable and conserved regions in the env gene
RA Willey R. W., Ruthledge R. A., Dias S., Folks T., Theodore T. S.,
RA Buckler C. E., Martin M. A.;
RT "Identification of conserved and divergent domains within the
RT envelope gene of the acquired immunodeficiency syndrome
RT retrovirus";
RL Proc. Natl. Acad. Sci. U. S. A. 83:5038-5042(1986).
- XX
- RN [21] art cds boundaries
RA Sodroski J. G., Goh W. C., Rosen C. A., Dayton A., Terwilliger E.,
RA Haseltine W.;
RT "A second post-transcriptional trans-activator gene required for
RT HTLV-III replication";
RL Nature 321:412-417(1986).
- XX
- CC Sequence for [7] was kindly supplied in computer readable form by

CC R. Crowley, 09/17/85. R. Patarca provided sites information and a
CC clean copy for [4], 09/16/85. Acquired immune deficiency syndrome
CC (AIDS) is caused by a retrovirus known by several names, perhaps
CC representing two separate strains: human T-cell lymphotropic
CC virus-III (HTLV-III), whose sequence is given below, and
CC lymphadenopathy-associated virus (LAV) are thought to be one strain
CC differing from AIDS-associated retrovirus type 2 (ARV-2) when
CC overall homology is the criterion. Some reading frame similarities
CC suggest that ARV-2 and LAV are more closely related. All three
CC viruses, whose sequences do not differ by more than 6%, are
CC believed to belong to the C type subfamily Lentiviridae, the "slow"
CC retroviruses. The BH10 sequence differs from BH8 and BH5 by 0.9% in
CC the coding regions and 1.8% in the noncoding regions, and the
CC authors of [2] believe that these are stable variants. The 5' and
CC 3' LTRs of BH10 and BH8 were not fully sequenced; the missing bases
CC (493-675 and 8608-9748) were filled in by [2] from the proviral
CC clone HXB2 [1]. The sequence below is that of BH10 with exception
CC of the variation at position 9197 which allows annotation of the
CC 27K coding sequence. The BH8 sequence spans bases 6033 to 9607, the
CC BH5 sequence spans bases 675 to 6038, and the HXB3 sequence [7]
CC spans bases 5778 to 8933. While this entry is offered as the
CC reference locus for the AIDS retroviral sequence loci, no claim is
CC being made that this sequence is more prevalent or typical than
CC others, all of which have been entered in this library with
CC annotation. The HTLV-III genome encodes at least six proteins or
CC polyproteins: gag, pol, env, TAT, 27K antigen and the sor 23K
CC product. The 3' ORF (positions 8797-9447) is truncated in BH10
CC (stop codon at positions 9196-9198), but reads through in BH8 and
CC other sequences to yield what is now called the 27K antigen. The
CC sequence below is from BH10 with exception of the variation at
position 9197 which allows annotation of the 27K coding sequence.
CC Additionally there are four short open reading frames, bases
CC 1248-1406, 4442-4542, 5592-5828 and 6095-6340, which are conserved
CC to a large degree. A seventh gene has been proposed based upon a
CC combination of mutational and regulatory evidence: called "ART" (for
CC anti-repression transactivator), its product appears to act
post-transcriptionally to relieve negative repression of gag and
CC env production [21]. The exon assignments for ART are putative, but
CC if they are corroborated, the ART protein would be 116 amino acids
CC in length. The mechanism for pol gene translation has not been
CC elucidated: a gag-pol fusion protein is possible; splicing or
CC frameshift have not been ruled out. The viral protease would be
CC determined by the region in question. Approximately two-thirds of
the variant sites in the gag and pol genes are "silent mutations",
CC while over half of those in the env gene are not. Reference [20]
CC defines divergent and conserved regions for the env gene. Because
CC of the excessive variability of the env gene, differences between
the sequences summarized herein and other env gene entries have not
CC been annotated; only HTLV-III sequence variations have been
CC included in the sites of this entry. Other entries will include
information for alignment with this entry, including the Zaire and
CC New York isolate sequences reported by [20]. The TAT protein
(trans-activator protein, approximately 14 kd) is an effector of an
CC autostimulatory pathway through interaction with a positive control
element, the trans-activating responsive sequence, TAR. TAT seems
CC to be a transcriptional control molecule in HTLV-I, but [11]
demonstrates that it is a post-transcriptional regulatory molecule
CC in HTLV-III. Deletion mutants in the TAT gene are incapable of
prolific replication and exhibit no cytopathic effects in T4+ cell
CC lines [19]. The TAR sequence(s) are found to be between -17 and +80
CC relative to the cap site +1 (base 455) and is highly conserved.
Enhancer sequences which need not be viral-specific are found
CC upstream from TAR [9],[11]. Three tandem decanucleotide Sp1 binding
sites are located between bases 377 and 408, of which site III
shows the strongest affinity for the cellular factor; intact, the
three sites cause up to a tenfold effect on transcriptional

CC efficiency in vitro ([17] (The authors demonstrate the existence of
 CC Sp1 in a human T-cell line). In addition to the ~9.4 kb genomic
 CC mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 have been
 CC detected. All are probably polyadenylated at the same site,
 CC position 8666 below, with a potential polyadenylation signal at
 CC 9642-8646, and capped at the same site, position 455, with a
 CC potential TATA box at 427-431. The doubly-spliced transcript of
 CC about 2.0 kb is responsible for the TAT message at least, and
 CC depending upon the acceptor site, also for the 50R and 27K
 CC messages, given that a single, albeit partial, mRNA exists for all
 CC three [18]. The acceptor splice for TAT is at position 5811 and the
 CC putative acceptor splice for 27K is at position 6010; the donor
 CC splice site in all three cases would be at position 6079 [18]. The
 CC doubly spliced message would also encode the newly proposed ART
 CC protein.

XX

FH	Key	From	To	Description
FT	RPT	1	634	5' LTR
FT	RPT	1	634	5' LTR
FT	VARIANT	82	82	a in BH10; g in H9
FT	VARIANT	101	101	g in BH10; a in H9
FT	VARIANT	108	108	a in [2]; H9; g in HXB2 [1]
FT	VARIANT	154	154	g in [2]; t in HXB2 [1], H9
FT	VARIANT	168	168	t in [2]; g in HXB2 [1], H9
FT	VARIANT	176	176	a in [2]; g in HXB2 [1], H9
FT	VARIANT	183	183	c in [2], H9; t in HXB2 [1]
FT	VARIANT	207	227	a in [2], H9; g in HXB2 [1]
FT	VARIANT	291	291	a in [2]; g in HXB2 [1], H9
FT	VARIANT	333	333	c in [2]; t in HXB2 [1], H9
FT	SITE	377	386	Sp1 binding site III [17]
FT	SITE	388	397	Sp1 binding site II [17]
FT	SITE	398	408	Sp1 binding site I [17]
FT	VARIANT	421	421	c in BH10, BH5; t in H9
FT	RPT	454	551	R repeat 5' copy
FT	RPT	454	551	R repeat 5' copy
FT	PROVRL	454	8666	HTLV3 virion RNA
FT	CAP	455	455	genomic mRNA start (cap site) [10]
FT	CAP	455	455	TAT, ART mRNA exon 1 start (cap site) [10], [18], [21]
FT	VARIANT	501	501	a in BH10, BH5, H9; g in HXB2 [1]
FT	SITE	636	653	primer (Lys-tRNA) binding site
FT	VARIANT	654	654	c in BH10, BH5; t in H9
FT	VARIANT	677	677	g in BH10, BH5; ggag in H9
FT	VARIANT	704	704	tga in BH10, H9; g in BH5 [2]
FT	CDS	787	2325	gag polyprotein precursor
FT	VARIANT	1290	1290	a in BH10; g in BH5 [2], H9
FT	VARIANT	1431	1431	a in BH10; g in BH5 [2], H9
FT	VARIANT	1455	1455	t in BH10, H9; c in BH5 [2]
FT	VARIANT	1611	1611	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1620	1620	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1656	1656	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1662	1662	t in BH10; c in BH5 [2], H9
FT	VARIANT	1673	1675	g in BH10, BH5; c in H9
FT	VARIANT	1722	1722	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1806	1806	g in BH10, BH5; a in H9
FT	VARIANT	1845	1845	a in BH10, BH5; g in H9
FT	VARIANT	1903	1903	a in BH10, H9; t in BH5 [2]
FT	VARIANT	1906	1906	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1923	1923	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1950	1950	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1953	1953	g in BH10, H9; t in BH5 [2]
FT	VARIANT	1958	1958	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1952	1952	c in BH10, H9; a in BH5 [2]
FT	VARIANT	2003	2003	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2013	2013	a in BH10, H9; a in BH5 [2]

FT	CDS	2381	5125	pol polyprotein (NH2-terminus uncertain: AA at 2381)
FT	VARIANT	2468	2468	g in BH10, BH5; a in H9
FT	VARIANT	2591	2591	c in BH10, H9; t in BH5 [2]
FT	VARIANT	2600	2600	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2741	2741	g in BH10; a in BH5 [2], H9
FT	VARIANT	2827	2827	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2858	2858	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2980	2980	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3007	3007	tta in BH10, H9; gtg in BH5 [2]
FT	VARIANT	3057	3057	a in BH10; g in BH5 [2], H9
FT	VARIANT	3122	3122	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3222	3222	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3302	3302	ag in BH10, H9; ga in BH5 [2]
FT	VARIANT	3366	3366	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3398	3398	g in BH10, BH5; a in H9
FT	VARIANT	3399	3395	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3755	3755	a in BH10, BH5; g in H9
FT	VARIANT	3767	3767	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3833	3833	t in BH10, BH5; c in H9
FT	VARIANT	3856	3856	t in BH10, BH5; c in H9
FT	VARIANT	3898	3898	c in BH10, BH5; t in H9
FT	VARIANT	3922	3922	a in BH10, H9; g in BH5 [2]
FT	VARIANT	3934	3934	a in BH10, BH5; g in H9
FT	VARIANT	3954	3954	g in BH10, BH5; c in H
FT	VARIANT	3962	3962	caa in BH10, H9; tag in BH5 [2]
FT	VARIANT	3977	3977	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3984	3984	c in BH10, H9; a in BH5 [2]
FT	VARIANT	3983	3993	a in BH10, H9; c in BH5 [2]
FT	VARIANT	4010	4010	a in BH10; g in BH5 [2], H9
FT	VARIANT	4016	4016	g in BH10, H9; a in BH5 [2]
FT	VARIANT	4029	4029	t in BH10, H9; c in BH5 [2]
FT	VARIANT	4049	4049	a in BH10; g in BH5 [2], H9
FT	VARIANT	4064	4064	c in BH10, H9; t in BH5 [2]
FT	VARIANT	4116	4116	a in BH10, BH5; c in H9
FT	VARIANT	4167	4167	g in BH10, BH5; c in H9
FT	VARIANT	4292	4292	t in BH10, H9; a in BH5 [2]
FT	CDS	5074	5652	sor 23K protein
FT	VARIANT	5156	5156	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5314	5314	t in BH10, BH5; c in H9
FT	VARIANT	5343	5348	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5401	5401	t in BH10, H9; c in BH5 [2]
FT	VARIANT	5412	5412	c in BH10, H9; t in BH5 [2]
FT	VARIANT	5549	5549	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5628	5628	g in BH10, H9; a in BH5 [2]
FT	VARIANT	5846	5846	g in BH10, H9, HXB3; a in BH5 [2]
FT	CDS	5864	6078	TAT protein, exon 2 (first expressed exon)
FT	VARIANT	5834	5834	a in BH10, H9, HXB3; c in BH5 [2]
FT	CDS	6003	6078	ART protein, exon 2 (first expressed exon; putative)
FT	VARIANT	6035	6045	cctccctcaagg in BH10, HXB3 [7]; gctcatcgaa in BH8 [2];
FT	VARIANT	6086	6086	g in BH5 [2], clone 12 cDNA [21]
FT	VARIANT	6096	6096	g in BH10, BH8, H9; a in HXB3 [7]
FT	VARIANT	6108	6108	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6113	6114	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6124	6124	gc in BH10, HXB3 [7], H9; gtaac in BH8 [2]
FT	VARIANT	6152	6152	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	CDS	6255	6325	g in BH10, HXB3 [7], BH8; c in H9
FT	VARIANT	6373	6373	envelope protein precursor (env)
FT	VARIANT	6474	6474	a in BH10, HXB3 [7], H9; t in BH8 [2]
FT	VARIANT	6748	6748	t in BH10, HXB3 [7], H9; g in HXB3 [7]
FT	VARIANT	6809	6929	t in BH10, HXB3 [7], H9; a in BH8 [2]
FT	VARIANT	6929		t in BH10, HXB3 [7], H9; c in BH8 [2]

FT	VARIANT	7088	7088	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7119	7119	a in BH10; HXB3 [7], H9; g in BH8 [2]
FT	VARIANT	7121	7123	cca in BH10, H9; cac in BH8 [2], HXB3 [7]
FT	VARIANT	7171	7172	gt in BH10, H9; aa in BH8 [2], HXB3[7]
FT	VARIANT	7187	7187	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7272	7273	aa in BH10, H9; gc in BH8[2], HXB3 [7]
FT	VARIANT	7291	7291	a in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	7343	7343	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7439	7454	gtttaatagtaacttgg in BH10, HXB3 [7], and H9
FT	VARIANT	7461	7461	a in BH10, BH8 [2]; g in HXB3 [7], H9
FT	VARIANT	7483	7483	c in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7521	7521	a in BH10, BH8 [2]; t in HXB3 [7], H9
FT	VARIANT	7574	7574	t in BH10, CH8 [2]; c in HXB3 [7], H9
FT	VARIANT	7636	7636	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7636	7637	cg in BH10, HXB3 [7], H9; gc in BH8[2]
FT	VARIANT	7645	7645	a in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8080	8081	ca in BH10, BH8 [2], H9; ac in H
FT	VARIANT	8127	8127	a in BH10, BH8 [2], H9; c in HXB[7]
FT	VARIANT	8131	8131	t in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	8135	8135	c in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8257	8257	g in BH10, BH8, HXB3; a in H9
FT	VARIANT	8273	8273	t in BH10, BH8, HXB3; g in H9
FT	VARIANT	8364	8364	g in BH10, HXB3 [7]; a in BH8 [2], H9
FT	CDS	8409	8454	TAT protein, exon 3 (AA at 8410)
FT	CDS	8409	8683	ART protein, exon 3 (putative; AA at 8411)
FT	VARIANT	8422	8422	t in BH10, HXB3 [7], clone 12 cDNA [21]; a in BH8 [2]; c in H9
FT	VARIANT	8484	8484	g in BH10, BH8, HXB3, clone 12 cDNA [21]; a in H9
FT	VARIANT	8657	8657	g in BH10, BH8 [2]; a in HXB3 [7], H9, clone 12 cDNA [21]
FT	VARIANT	8672	8672	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT	VARIANT	8692	8692	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT	VARIANT	8748	8748	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8759	8758	g in BH10, H9; c in BH8 [2]; a in HXB3 [7], clone 12 cDNA [21]
FT	VARIANT	8771	8771	t in BH10, HXB3 [7], clone 12 cDNA [21], H9; c in BH8 [2]
FT	CDS	8827	9447	27K protein, exon 3 (first expressed exon)
FT	VARIANT	8857	8857	g in BH10, BH8, HXB3, clone 12 cDNA [21]; a in H9
FT	VARIANT	8824	8824	c in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8867	8867	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	8978	8978	a in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT	VARIANT	8985	8985	t in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT	VARIANT	8987	8987	a in BH10, BH8; c in H9, clone 12 cDNA [21]
FT	VARIANT	8994	8994	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	9019	9019	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT	RPT	9116	9748	3' LTR
FT	VARIANT	9183	9186	t in BH10, clone 12 cDNA [21]; c in BH8 [2]
FT	VARIANT	9197	9197	a in BH8 [2], H9, clone 12 cDNA [21];

FT	VARIANT	9218	9218	a in BH10 [2]
FT	VARIANT	9219	9223	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT	VARIANT	9273	9279	ga in BH10, clone 12 cDNA [21], H9; ag in BH8[2]
FT	VARIANT	9283	9283	g in BH10, BH8, clone 12 cDNA [21]; t in H9
FT	VARIANT	9284	9284	t in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	VARIANT	9291	9291	t in BH10, H9, clone 12 cDNA [21]; a in BH8 [2]
FT	VARIANT	9297	9297	a in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	VARIANT	9354	9354	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT	VARIANT	9406	9406	g in BH10, HIVDSM>, H9; t in BH8 [2]
FT	VARIANT	9448	9448	a in BH10, BH8; g in H9, clone 12 cDNA [21]
FT	VARIANT	9533	9583	c in BH10; t in BH8 [2], H9, clone 12 cDNA
FT	RPT	9570	9666	c in BH10, BH8, clone 12 cDNA [21]; g in H9
FT	VARIANT	9616	9616	R repeat 3' copy
FT	VARIANT	9621	9621	g in HXB2; a in H9, clone 12 cDNA [21]
FT	VARIANT	9663	9663	g in HXB2; a in H9, clone 12 cDNA [21]
FT	POLYA	9666	9666	t in BH10, H9; tg in clone 12 cDNA [21]
FT	POLYA	9666	9666	TAT, ART, 27K mRNA exon 3 end (poly-A site) [10],[18],[21]
FT	POLYA	9666	9666	genomic mRNA end (poly-A site) [10]

XX
 SQ Sequence 3749 BP; 3431 A; 1781 C; 2369 G; 2168 T; 0 other;
 Initial Score = 1872 Optimized Score = 2176 Significance = 0.00
 Residue Identity = 89% Matches = 2243 Mismatches = 150
 Gaps = 106 Conservative Substitutions = 0

'70 80 90 100 110 120
 GATAATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GGTTGAGATGGTGCAACCATGCTCTTGCGATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAG
 6300 6310 6320 6330 6340 6350 6360

```

  130          140          150          160          170          180
-CT-----GTGTGCAA-----GAAAGCAA-CCACCA-CTCTATTTGTGCATCAGATGCTAAAGCATATGAT
   :       : : : :       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TCTATTATGGGGTACCTGTGTGGAGGAAGCAACCACCACTCTATTTC-----TGC---ATCAGATGCT
6370      6380      6390      6400      6410      6420

```

130	200	210	220	230	240	
A---CAGAGG-TACATA-----AT--GTTTGGGCCACACATGCCTG--TGTACCCACAGA-CCCCAACCCAC	:	:	:	:	:	
AAAGICATAAGATACAGAGGTACATAATGTTGGGC-CACA--CATGCCTGTGTACCCACAGACCCCCAAC-C						
6430	6440	6450	6460	6470	6480	6490

1100 1110 1120 1130 1140 1150 1160
 TCTT TAAAGCAATCTGAGGAGGGAAACCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TCTT TAAAGCAATCTGAGGAGGGAAACCCAGAAATTGTAACGCACAGTTTAATTGTGGAGGGGAATTTTCT
 7340 7350 7360 7370 7380 7390 7400

 1170 1180 1190 1200 1210 1220 1230
 ACTGTAAT TCGAAGACACTGTTAATAGTACTTGGAGTACTGAAAGGGTCAAATAACA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 ACTGTAATTCAGACACTGTTAATAGTACTTGGAGTACTGAAAGGGTCAAATAACA
 7410 7420 7430 7440 7450 7460 7470

 1240 1250 1260 1270 1280 1290 1300
 CTGAAAGGAACTGACACAAATCACACTCCCATGCAGAAATAAACATGTGGCAGGAAGTAGGAA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CTGAAAGGAACTGACACAAATCACACTCCCATGCAGAAATAAACATGTGGCAGGAAGTAGGAA
 7480 7490 7500 7510 7520 7530 7540

 1310 1320 1330 1340 1350 1360 1370 1380
 AAGCATACTATGCCATCCCATCAGCGGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AAGCATACTATGCCATCAGCGGGACAAATTAGATGTTCATCAAATATTACAGGGCTGCTATTAACAA
 7550 7560 7570 7580 7590 7600 7610 7620

 1390 1400 1410 1420 1430 1440 1450
 GAGAATGCTTAAACAAACAAATGGTUCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GAGAATGCTTAAACAAACAAATGGTUCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
 7630 7640 7650 7660 7670 7680 7690

 1460 1470 1480 1490 1500 1510 1520
 GAAGTGAAATTATATAATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GAAGTGAAATTATATAATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA
 7700 7710 7720 7730 7740 7750 7760

 1530 1540 1550 1560 1570 1580 1590
 GAGTGCTGAGGAGAAAGAGCAGTGGGAATAGGAGCTTGTTCCTGGGTTCTGGGAGCAGCAGGAA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GAGTGCTGAGGAGAAAGAGCAGTGGGAATAGGAGCTTGTTCCTGGGTTCTGGGAGCAGCAGGAA
 7770 7780 7790 7800 7810 7820 7830

 1600 1610 1620 1630 1640 1650 1660
 AGACCTATGAGGCACTGCAATGAGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GCACCTATGAGGCACTGCAATGAGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
 7840 7850 7860 7870 7880 7890 7900

 1670 1680 1690 1700 1710 1720 1730 1740
 AGAAGCAATTGCTGAGGGCTATTGAGGCGCAACAGCAGTCTGTTGCAACTCACAGTCTGGGAGCATCAAGCAGC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 AGAAGCAATTGCTGAGGGCTATTGAGGCGCAACAGCAGTCTGTTGCAACTCACAGTCTGGGAGCATCAAGCAGC
 7910 7920 7930 7940 7950 7960 7970 7980

 1750 1760 1770 1780 1790 1800 1810
 TCCAGGCAAGAACTGGCTTGGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 TCCAGGCAAGAACTGGCTTGGGAAAGATAACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTG
 7990 8000 8010 8020 8030 8040 8050

 1820 1830 1840 1850 1860 1870 1880
 GAAAACGTTTGGCAAGACTGCTTGGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTTGGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GAAAACGTTTGGCAAGACTGCTTGGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTTGGA
 8060 8070 8080 8090 8100 8110 8120

1860 1880 1900 1920 1930 1940 1950
 ATATCATORCTTACATTCAGTGAGGACAGAGAAATTAAACAATTACACAAGCTTAATACTACATTCCCTTAATTGAAG
 : : : : : : : :
 ATATCATORCTTACATTCAGTGAGGACAGAGAAATTAAACAATTACACAAGCTTAATACTACATTCCCTTAATTGAAG
 8130 8140 8150 8160 8170 8180 8190
 1960 1970 1980 1990 2000 2010 2020
 AATCGCAAAACGGACGGAAAGCATGAAACAAGAATTATTGGATTAGATAAAATGGGCAAGTTGTGGAAATT
 : : : : : : : :
 AATCGCAAAACGGACGGAAAGCATGAAACAAGAATTATTGGATTAGATAAAATGGGCAAGTTGTGGAAATT
 8200 8210 8220 8230 8240 8250 8260
 2030 2040 2050 2060 2070 2080 2090 2100
 GGTTCAGCTATTAACAACTTGCCTCTTGTATATAAAATTTCATAATGATAGTAGGAGGCTTGGTAGGTTAA
 : : : : : : : :
 GGTTCAGCTATTAACAACTTGCCTCTTGTATATAAAATTTCATAATGATAGTAGGAGGCTTGGTAGGTTAA
 8270 8280 8290 8300 8310 8320 8330 8340
 2110 2120 2130 2140 2150 2160 2170
 AGATAGTTTTCTCTTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCC
 : : : : : : : :
 GAATAGTTTTCTCTTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCC
 8350 8360 8370 8380 8390 8400 8410
 2180 2190 2200 2210 2220 2230 2240
 ACCTTCCCACACGGACGGACCCACAGCCGAAAGGAATAGAAGAAGAAGGTGGAGAGAGACAGAGACA
 : : : : : : : :
 ACCTTCCCACACGGACGGACCCACAGCCGAAAGGAATAGAAGAAGAAGGTGGAGAGAGACAGAGACA
 8420 8430 8440 8450 8460 8470 8480
 2250 2260 2270 2280 2290 2300 2310
 GATGCGATTTGGATTAACGAGGCTTCTGGACACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGC
 : : : : : : : :
 GATGCGATTTGGATTAACGAGGCTTCTGGACACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGC
 8490 8500 8510 8520 8530 8540 8550
 2320 2330 2340 2350 2360 2370 2380
 TAQIACCGCTTACGAGGAACTTCTTGTGGAAACTTCTGGGACGCAGGGGGTGGGAA
 : : : : : : : :
 TAQIACCGCTTACGAGGAACTTCTTGTGGAAACTTCTGGGACGCAGGGGGTGGGAA
 8560 8570 8580 8590 8600 8610 8620
 2390 2400 2410 2420 2430 2440
 GCCTTCATATTTGTTGAACTTCTACAGTATTGGAGTCAGGAACCTAAAG
 : : : : : : : :
 GCCTTCATATTTGTTGAACTTCTACAGTATTGGAGTCAGGAACCTAAAG
 8630 8640 8650 8660 8670 X

3. KUNZ-158-10301-529

HIVMAL05 Human lymphadenopathy virus (MAL isolate), complet

ID HIVMAL05 standard RNA: 9229 BP.
 XX
 AC M03415;
 XX
 DT 07-JULY-1987 (Incorporated)
 XX
 DE Human lymphadenopathy virus (MAL isolate), complete genome.
 XX
 KW acquired immunodeficiency syndrome; env gene; gag gene; genome;
 KW long terminal repeat; pol gene; polyprotein; provirus;
 KW provirus; structural genes
 XX
 DS Human lymphadenopathy virus

OC Mammal-specific enveloped viruses; Retroviridae.

XX E11 (Access 1-E1285)

RN Allizon M., Meir-Hubron S., Montagnier L., Sonigo P.;
RT "Genetic variability of the AIDS virus: Nucleotide sequence
analysis of two isolates from African patients";
RT Cell 49(2):741 (1986).

XX Acquired immune deficiency syndrome (AIDS) is caused by a
retrovirus known by several different names, probably representing
two separate strains: human T-cell lymphotropic virus-III
(HTLV-III) and lymphadenopathy-associated virus (LAV) are thought
to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the
other. All three viruses, whose sequences do not differ by more
than about 8%, are believed to belong to the retroviral subfamily
Lentiviridae, or "slow" viruses. For the details of the annotation
and for other pertinent references, see the HIV reference entry.

XX

FH	Key	From	To	Description
FT	RPT	1	95	R repeat 5' copy
FT	RPT	1	177	5'LTR
FT	SITE	178	193	primer (Lys-tRNA) binding site
FT	CDS	350	1057	gag polyprotein
FT	CDS	1863	4871	pol polyprotein (NH2-terminus uncertain; AA at 1863)
FT	CDS	4216	5154	sor 23K protein
FT	CDS	5154	5424	urfC
FT	CDS	5405	5619	tat protein, exon 2 (first expressed exon)
FT	CDS	5713	6374	envelope polyprotein precursor
FT	CDS	7959	9007	tat protein, exon 3 (AA at 7960)
FT	CDS	8390	8608	27K protein
FT	CDS	8678	8920	3'LTR
FT	RPT	9154	9224	R repeat 3' copy

XX SQ Sequence: 3209 SP/ 3255 AV/ 1627 C/ 2204 G/ 2043 T/ 0 other;

Initial Score = 916 Optimized Score = 2041 Significance = 0.00
Residue Identity = 84% Matches = 2066 Mismatches = 349
Gaps = 43 Conservative Substitutions = 0

X 10 20 30 40 50 60 70
ATGAGAATGAACTTGTACATGACTTGTTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATGAGAATGAACTTGTACATGACTTGTTGGAGATGGGGCATGATGCTCCTTGGGATG
5800 5810 5820 5830 5840 5850 5860

80 90 100 110 120 130 140
TTGGAGATGAACTTGTACATGACTTGTTGGAGATGGGGTACCTGTGTGGAAAGGAAGCA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TTGGAGATGAACTTGTACATGACTTGTTGGAGATGGGGTACCTGTGTGGAAAGGAAGCA
5870 5880 5890 5900 5910 5920 5930

150 160 170 180 190 200 210
ACGGAGATGAACTTGTACATGACTTGTTGGAGATGGGGTACCTGTGTGGAAAGGAAGCA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ACGGAGATGAACTTGTACATGACTTGTTGGAGATGGGGTACCTGTGTGGAAAGGAAGCA
5940 5950 5960 5970 5980 5990 6000 6010

220 230 240 250 260 270 280
GGGGTACCTGTGTGGAAAGGAAGCA
||||| ||||| ||||| ||||| ||||| ||||| |||||
GGGGTACCTGTGTGGAAAGGAAGCA
6020 6030 6040 6050 6060 6070 6080

290 300 310 320 330 340 350 360
 AACAAATGAACTTGTAAACGATTCATCGAGCATATAATCAGTTATGGGATCAAAGCCTAAAGCCATGTGTA
 TTT
 AAAAAATGACATTTTGCGAGATTCATCGAGCATATAATCAGTTATGGGATCAAAGCCTAAAACCATGTGTA
 6090 6100 6110 6120 6130 6140 6150
 370 380 390 400 410 420
 AAGTAAAGCAUAGTTCATTTTAAAGTCACGTATTG---GGGAATGCTACTAAT---ACCAACT
 TTT
 AACCTAACCGCTCTTGTGAACTGACTAATGTGAATGGGACTGCTGTGAATGGGACTAATGCT
 6160 6170 6180 6190 6200 6210 6220
 430 440 450 460 470 480 490
 AGTGTAGGCGTAACTGCTCTTGTGAACTGACTAATGTGAATGGGACTGCTCTTCAATATC
 TTT
 GGGAGT---AGAAAGACTAATGCAAAATTGAAATGGAAATTGGAGAAGTGAAACTGCTCTTCAATATA
 6230 6240 6250 6260 6270 6280 6290
 500 510 520 530 540 550 560
 AGCGAGGAACTAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 TTT
 ACCCGATGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 6300 6310 6320 6330 6340 6350 6360
 570 580 590 600 610 620 630
 GATGTTTAACTACGAACTATGCTTGACAAATTGTAAACACCTCAGTCATTACACAGGCCTGTC
 TTT
 GAT
 6370 6380 6390 6400 6410 6420 6430
 640 650 660 670 680 690 700 710
 TGTTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 TTT
 ACCGTTGGTAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 6440 6450 6460 6470 6480 6490 6500
 720 730 740 750 760 770 780
 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 GTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 6510 6520 6530 6540 6550 6560 6570 6580
 790 800 810 820 830 840 850
 ACTGAACTACGTTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 TTT
 ACTGAACTACGTTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 6650 6660 6670 6680 6690 6700 6710 6720
 860 870 880 890 900 910 920
 GCTGAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
 TTT
 ACTGAACTACGTTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACTTGTGAACT
 6730 6740 6750 6760 6770 6780 6790
 930 940 950 960 970 980 990
 GAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 TTT
 AGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 6770 6780 6790
 1000 1010 1020 1030 1040 1050 1060 1070
 GCGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 TTT
 GCGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
 6800 6810 6820 6830 6840 6850 6860

1050 1060 1070 1080 1090 1100 1110 1120 1130 1140
 TTTTGGGTTCTTGTGAACTATTTTACGCACTTCAGGGGGACCCAGAAATTGTAACGCACAGTTT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTTTGTGAACTTCAGGGGGACCCAGAAATTACAACACACAGTTT
 6870 6880 6890 6900 6910 6920 6930 6940

 1150 1160 1170 1180 1190 1200 1210
 GATTCCTTGGCGCTTGTGAACTTCAGGGGGACCCAGAAATTACAACACACAGTTT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AATTTGAACTTCAGGGGGACCCAGAAATTACAACACACAGTTT
 6950 6960 6970 6980 6990 7000 7010

 1220 1230 1240 1250 1260 1270 1280
 AGTGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTACAACACTCCCAGCAGAATAAAAACAAATTATA
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AATTTGAACTTCAGGGGGACCCAGAAATTACAACACTCCCAGCAGAATAAAAACAAATTATA
 7020 7030 7040 7050 7060 7070 7080

 1290 1300 1310 1320 1330 1340 1350
 AACATGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTAGATGTTCATCAAAT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AATTTGAACTTCAGGGGGACCCAGAAATTAGATGTTCATCAAAT
 7090 7100 7110 7120 7130 7140 7150

 1360 1370 1380 1390 1400 1410 1420
 ATTACAGGAACTTCAGGGGGACCCAGAAATTAGATGACAATAGTGAACATGAGACCTTAAGACCT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ATTACAGGAACTTCAGGGGGACCCAGAAATTAGTGAACATAGTGAACATGAGACCTTAAGACCT
 7160 7170 7180 7190 7200 7210 7220

 1430 1440 1450 1460 1470 1480 1490
 GGAGGGGGTGTGAACTTCAGGGGGACCCAGAAATTATAAAATATAAAAGTAGTAGAAAAATTGAACCATTA
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GGAGGGGGACCCAGAAATTATAAAATATAAAAGTAGTAGTAAGAATTGAACCCCTA
 7230 7240 7250 7260 7270 7280 7290

 1500 1510 1520 1530 1540 1550 1560
 GGAGGGGGTGTGAACTTCAGGGGGACCCAGAAATTAGGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTG
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 GGAGGGGGACCCAGAAATTAGGAGAGGGGGAGAAAAAGAGCAATAGGAGCTAGGAGCCATG
 7300 7310 7320 7330 7340 7350 7360

 1570 1580 1590 1600 1610 1620 1630 1640
 TTGTGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCGACGGTCAATGACGCTGACGGTACAGGCCAGA
 ||| ||| ||| ||| ||| ||| ||| |||
 TTGTGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCGACGGTCACTAACGCTGACGGTACAGGCCAGA
 7380 7390 7400 7410 7420 7430 7440

 1650 1660 1670 1680 1690 1700 1710
 CAATTTGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 ||| ||| ||| ||| ||| ||| |||
 CAATTTGTTGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 7450 7460 7470 7480 7490 7500 7510

 1720 1730 1740 1750 1760 1770 1780
 CGAGGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 ||| ||| ||| ||| ||| |||
 CGAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 7520 7530 7540 7550 7560 7570 7580

 1790 1800 1810 1820 1830 1840 1850
 CGAGGGGGTGTGAACTTCAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 ||| ||| ||| ||| ||| |||
 CGAGGGGGACCCAGAAATTGGCTGAGGGCTATTGAGGCGCAACAGCATCTGTG
 7590 7600 7610 7620 7630 7640 7650

4. KUNZ-158-CL33. SEQ

HIVELICG Human lymphadenopathy virus (ELI isolate), complete genome

ID HIVELOC standard FNA 9176 BP.

xx

AC

XX

DT

Human lymphadenopathy virus (EL-1 isolate): complete genome

1

KW acquired immune deficiency syndrome; env gene; gag gene; genome;
KW long terminal repeat; pol gene; polyprotein; provirus;
KW reverse transcriptase.

OS Human lymphadenopathy virus
OC *Viridae*: ss-RNA enveloped viruses; *Retroviridae*.

XX
RN [1] (bases 1-8176)
RA Alizon M., Watrin-Holzman S., Montagnier L., Sonigo P.;
RT "Genetic variability of the AIDS virus: Nucleotide sequence
RT analysis of two isolates from African patients";
RL Cell 40:67-74 (1985).

XX
CC Acquired immune deficiency syndrome (AIDS) is caused by a
CC retrovirus known by several different names, probably representing
CC two separate strains: human T-cell lymphotropic virus-III
CC (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought
CC to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the
CC other. All three viruses, whose sequences do not differ by more
CC than about 5%, are believed to belong to the retroviral subfamily
CC Lentiviridae, or "slow" viruses. For the details of the annotation
CC and for other pertinent references, see the HIV reference entry.

XX				
FH	Key	From	To	Description
FH				
FT	RPT	1	83	R repeat 5' copy
FT	RPT	1	180	5' LTR
FT	SITE	182	188	primer (Lys-tRNA)
FT	CDS	336	1230	gag polyprotein
FT	CDS	1231	4842	pol polyprotein
FT				uncertain: AA at
FT	CDS	4537	5165	end 23K protein
FT	CDS	5166	5235	orfC
FT	CDS	5236	5252	orfD
FT	CDS	5770	6171	envelope polypro
FT	CDS	6333	6513	27K protein
FT	RPT	6825	8178	3' LTR
FT	RPT	8070	9178	R repeat 3' copy

SQ Sequence: 11 46 EPI: 7733 A; 1632 C; 2179 G; 2032 T; 0 other;

Initial Score = 1246 Optimized Score = 1893 Significance = 0.00
 Residue Identity = 73% Matches = 1959 Mismatches = 435
 Gaps = 103 Conservative Substitutions = 0

70 100 90 100 110 120
-CCCTTG-GCATATG---ATGAT-UT-GTAGTCTACAGAAAAATTGT-GGGTCACAGTCTATTATGGGGTAC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATCTGTTG-CGATG-TGCGATGTTGCTGCTTGATATGATGACCTGTAGTGCTGCAGACAATCTGTGGGTCA
1810 5770 5830 5840 5850 5860 5870

130	135	140	145	150	155	160	165	170	175	180	185	190
CCTGTATGAGTACATGATGATACTTGTATTTTG-TGCATCAGATGCTAAAGCATATGATA	T	A	T	T	T	T	T	T	T	T	T	T
GAT	G	A	T	A	T	G	A	T	G	A	T	G
5930	5935	5940	5945	5950	5955	5960	5965	5970	5975	5980	5985	5990

260 260 260 260 260 260 260
TGTATGTTTCTTGACAGAATTCCTCTGTAAACCACAGACC--CCAACCCAC-AAGAAGTAGTATTGGTA

6770 6771 6772 6773 6774 6775 6776 6777
 1030 1040 1050 1060 1070 1080 1090
 TTTTAACTGGGTTTATTAACGAACTAATAGGACAAACACATTGTAATATTAGTAGAGCACAAT
 6778 6779 6780 6781 6782 6783 6784 6785
 1100 1110 1120 1130 1140 1150 1160 1170
 AGGAGATCCCTGGGAACTTAAACGAACTAATAGGACAACTTAAATTGTGGAGGGGAATTCTTCACTGTA
 6786 6787 6788 6789 6790 6791 6792 6793
 1180 1190 1200 1210 1220 1230 1240
 GGTGAGACATTTTATAATTTACTTGGTTAACAGTACTTGGA---AGTACTGAAGGGTCAAATAACACTG
 6794 6795 6796 6797 6798 6799 6800 6801
 1250 1260 1270 1280 1290 1300 1310
 AGAGGAGTTCCTGGGAACTTAAACGAACTAATAGGACAACTTAAACAT-GTGGCAGGAAGTAGGAAA
 6802 6803 6804 6805 6806 6807 6808 6809
 1320 1330 1340 1350 1360 1370 1380
 AGCCTTTATGCGGTTTAACTGGGAACTTAAACGAACTTAAACGAACTTAAAGATGGTGGCAG---GCAGGAA-A
 6810 6811 6812 6813 6814 6815 6816 6817
 1390 1400 1410 1420 1430 1440 1450
 GATGTTGGTTCCTGGGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAA
 6818 6819 6820 6821 6822 6823 6824 6825
 1460 1470 1480 1490 1500 1510 1520
 AGTCAATTGTTGAGTATAATGTAACAAATTGAAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGA
 6826 6827 6828 6829 6830 6831 6832 6833
 1530 1540 1550 1560 1570 1580 1590
 ATGTTGGGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAA
 6834 6835 6836 6837 6838 6839 6840 6841
 1600 1610 1620 1630 1640 1650 1660 1670
 AGTATGGGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAA
 6842 6843 6844 6845 6846 6847 6848 6849
 1680 1690 1700 1710 1720 1730 1740
 AGCAGATTGTTGAGGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAA
 6850 6851 6852 6853 6854 6855 6856 6857
 1750 1760 1770 1780 1790 1800 1810
 GAGGAGGAGTTCCTGGGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAACTTAAACGAA

CATGCCGACATTTCTTGTGAGATAACCTAAAGGATCAACAGCTCCTAGGAATTGGGGTTGCTCTGGA
 7430 7500 7510 7520 7530 7540 7550 7560
 1880 1880 1890 1890 1890 1870 1880
 AAACUCAATTTCGGCGCTTGCGATGCTTAGTTGGAGTAATAAATCTCTGGAACAGATTGGAAAT
 :
 AAGATGCTTGCCTGCTTAAATGCGCTGAAAGTCTAETTGGAGTAATAGATCTCTAAATGAGATTGGCAG
 7570 7580 7590 7600 7610 7620 7630
 1890 1890 1910 1920 1930 1940 1950
 AAGAAGGTTGATGATGATGAGAAATTAACATTACACAAGCTTAATACATTCCCTTAATTGAAGAA
 :
 AACATGAACTTGGTGTGATGAGAAATTAACATTACACAGGCTTAATATAGCTTAATTGAGGAA
 7640 7650 7660 7670 7680 7690 7700
 1960 1970 1980 1990 2000 2010 2020 2030
 TGTGAGATGCGCGAAATGAAACAGAAATTATTGGAAATTAGATAATGGCAAGTTGTGGAAATTGG
 :
 TGTGAGATGCGCGAAATGAAACAGAAATTGGAAATTGGACAAAGTGGCAAGTTGTGGAAATTGG
 7710 7720 7730 7740 7750 7760 7770
 2040 2050 2060 2070 2080 2090 2100
 TTTCAGATTTGAAATTTGGCTTGCTTGGTATATAAAAAATATTCTATAATGATAAGTAGGAGGGCTTGGTAGGTTAAGA
 :
 TTTCAGATTTGAAATTTGGCTTGCTTGGTATATAAAAAATATTCTATAATGATAATAGGAGGGCTTGTAGGTTAAGA
 7780 7790 7800 7810 7820 7830 7840
 2110 2120 2130 2140 2150 2160 2170
 ATACCTTTCGTTAGCTTCTCTTAAATAGAGTTAGGCAGGGATATTCAACCATTATCGTTTCAGACCCAC
 :
 ATACCTTTCGTTAGCTTAAATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTTCAGACCCCTC
 7850 7860 7870 7880 7890 7900 7910 7920
 2180 2190 2200 2210 2220 2230 2240
 CTCAGAAAGTGGCGACCGCGGCGAGGAATAGAAGAAAGGTTGGAGAGAGACAGAGACAGA
 :
 CTCAGAAAGTGGCGACCGCGGCGAGGAACACAAAGAAGAAGGTTGGAGAGAGCGAGGCAGAGACAGA
 7930 7940 7950 7960 7970 7980 7990
 2350 2360 2370 2380 2390 2300 2310
 TGGCAATTGGATTTGCGGGCGCTTCTGGGACGATCTGGGAGGCCTTGTGCCTCTTCAGCTA
 :
 TGGCAATTGGATTTGCGGGCGCTTCTGGGACGCTGCGGAGGCC-TGTGCCTCTTCAGCTA
 8000 8010 8020 8030 8040 8050 8060
 2320 2330 2340 2350 2360 2370 2380 2390
 CGCGCGCTTGGCTTAACTTACCTTCTGATTAACGAGGATTGTGGAAACTTCTGGGACGCAAGGGGGTGGGAAGC
 :
 CGCGCGCTTGGCTTAACTTACCTTCTGATTAACGAGGATTGTGGAAACTTCTGGGACGCAAGGGGGTGGGAACAT
 8070 8080 8090 8100 8110 8120 8130
 2400 2410 2420 2430 2440
 CCTCGAGATTTGCGGGCGCTTCTGGGACGCTTCTGGGAGTCAGGAACCTAAG
 :
 CCTCGAGATTTGCGGGCGCTTCTGGGACGCTTCTGGGAGTCAGGAACCTGAGG
 8140 8150 8160 8170 8180 X

5. KUNZ-158-GL33 508

RIGHTLV4C STMV-1 (HTLV-1) partial provirus DNA (lambda PKE10)

ID: HTLV-1C Standard: DNA: 5391 BP.

XX

AC: KX63911; MM02534;

XX

DT: 24 MAR 1985 (Annotation)

1000 1050 1100 1110 1120 1130
 200 250 300 310 320 330
 -GAAATTCTTACATGTTGGAAAA--ATGA-CAT--GGTAGAAC---AGATGCATGAGGATATAATCAG-TT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GATTAATTCGGTTCTTATCCATTATGAGATGCAATAAAAGTGAGACAGATA
 1150 1160 1170 1180 1190 1200
 240 250 260 270 280 290
 TAT-GGGAA---CA-AGCCCTAAA-GCCATGTGTAACCCCC-AC-TCTGTGTTAGT-TTAAAGTGCA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GATGGAAATTGAAATCAAAUACAAATAACACAGCACCAACATCAGCACCAAGTATCAGAAAAAA
 1210 1220 1230 1240 1250 1260 1270
 300 310 320 330 340 350 360 370 380 390
 CTATTTGGGAAATGCTACTAATACCAATACCTAG-TAATACCAATAGTAGTAGTAGCGGGGAA-ATGA-TGATGG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TAATGATGTTGAAATGAGCTAGTTCTTGTA-TAGCTCAGAATAATTGCACAGGCTTGGAACAGAGCAAATG
 1280 1290 1300 1310 1320 1330 1340 1350
 400 410 420 430 440 450 460
 AGAAGGAAATGAAATGAAACTGCTCTTCAATATCAGCACAGNATAAGAGGTAAAGGGTGCAGAAAG--AATA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ATATGCTGAAATGAACTGAGGTTAAAHAGAG-ACAAGACAAAGGGTACAATGAAACTTGGTACTC
 1360 1370 1380 1390 1400 1410 1420
 470 480 490 500 510 520 530
 TAGGAAATGTTTGTGAAAC-AAGGAAATAGCACT-GATAATGAAAGCAGATGC--TACATAAATCA-CTG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 1430 1440 1450 1460 1470 1480
 540 550 560 570 580 590 600
 TGC--ATTT-TTCTATAAACTGCTATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGTTG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TAGAGATTGTTGTTGAAAC-AAGGAAATAGCACT-GATAATGAAAGCAGATGC--TACATAAATCA-CTG
 1490 1500 1510 1520 1530 1540 1550
 610 620 630 640 650 660 670
 TAACACCTGAACTCATTACACAGGCTGT-CCAGGGTATCCTTGAGCCAATTCCCACATTATTGTGCC
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TAACACTTCTGTTTATCCAAAGAGCTTGACAAACAT-TATTGGGATACTATTAGATTAGGTATTGTGCAC
 1450 1500 1510 1520 1530 1540 1550
 680 690 700 710 720 730 740
 CGGTGGTTTACGATCTAAATGTAATAATA-AGACGTTCAATGGAACAGGGAC--CATGTACAAATGT-C
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CTCAGGGTTATGTTGCTTGTAGAATGACACAAATTATTCAAGGCTTATGCCTAAATGTTCTAAGGTGG
 1560 1570 1580 1590 1600 1610 1620 1630
 750 760 770 780 790 800
 AGCACAGTACAACTGACACATGGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TGGCTCTCTTC-ATG--CACAAGGATGAGGAGACACAG-ACT-TCTACTTGGTTGGCTTAAATGGAACTAG
 1640 1650 1660 1670 1680 1690
 810 820 830 840 850 860 870
 AGCAAGAAAGT-AGAGGTAGTAAATTAGATCTGCCAATTTCACAGACAATGCTAAACCATAA-TAG--TACAGC
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AGCAAGAAATGAAACTTATATTAA-CTGGC-ATGGTAGGGATAA---TAGGACTATAATTAGTTAAATA
 1700 1710 1720 1730 1740 1750 1760
 880 890 900 910 920 930 940
 TGAACCAAATCTGTAGAAATTAATTGACAGACCAACAAACATACAAGAAAAGTATCCGTATCCAGAGGG
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AGTATTATGAAATCTAACAAATGAGAGA-CGAGGA--AAT--AAG--ACAGTTT---TA-CCA---GT
 1770 1780 1790 1800 1810 1820
 950 960 970 980 990 1000 1010
 GACCA--GGGAGGCGATT--TGTACAATAGGAAAGGAAATGAGAC-AAGCA--CATTTGAAACATT
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GACGATTATGCTGGATTGGTTTCCACTCACACCA-ATCAATGATAGGCCAAAGCAGGCATGGT---GTT

2500	2550	2600	2610	2620	2630	2640	2650
1780	1790	1800	1810	1820	1830	1840	
ACCTAAAGGATCAACAGCTTGTGGATTGGGGTTGCTCTGGAAAAACTCATTGCACCACTGCTGTGCC	TT						
ACTTAAAGGATCAACAGCTTGTGGATTGGGGTTGCTCTGGAAAAACTCATTGCACCACTGCTGTGCC	TT						
2660	2670	2680	2690	2700	2710	2720	
1850	1860	1870	1880	1890	1900	1910	
GG----AATCTAATTGGAGTAATAAACTCTGGAACAGATTGGAAATAACATGACCTGGATGGAGTGGGACA	TT						
GCGCAAAATGCAAGT----CTAACACCAGACTGGAAACA-AT---GA-TA-----CTTGCGCAAGAGTGGGAGC	2730	2740	2750	2760	2770		
1920	1930	1940	1950	1960	1970	1980	
GAGAAATTAAAC-AATTACACAAAGCTTAATACATTCCTTAATTGAAGAAATCGCAAAACCGAGCAAGAAAAGAAT	TT						
GAAGAAGTTGACTTCTTGGAGGAAAATATAACA-GCCCTCCTAGAAGAGGGCACAAATTCAACAAAGAGAGAAC	2780	2790	2800	2810	2820	2830	2840
2850							
1990	2000	2010	2020	2030	2040	2050	
GAACAAGAATTATTGGATTAGATAAAATGGGCAAGTITGTGG-AATTGGTTAACATAACAAATTGGCTGTG	TT						
ATG1ATGAATTACAAAAGTTAAATAGCTGGG-ATGTETTTGGCAATTGGTTGACCTTGCTTCTGGATAAA	2860	2870	2880	2890	2900	2910	2920
2060	2070	2080	2090	2100	2110	2120	
GTA1ATAAAATTATTCA--TAATGATAGTAGGAGGCTTGGTAGGTTAAGAATAGTTTGCTGTACTTTCT	TT						
GTA1AT-ACAATATGGAATTATG-TAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAA	2930	2940	2950	2960	2970	2980	2990
2130	2140	2150	2160	2170	2180		
ATACGTGAATAGAGTT-AGGCAGGGATAATTACCA---TTATCGTTTCAGACCCAC---CTCCCA---AC--	TT	TT	TT	TT	TT		
ATACGTAGCTA-ATTTAGGCAGGGTATADGCCAGTGTTC-TCACCCACCCCTTTATTCAGTAGACTC	3000	3010	3020	3030	3040	3050	3060
3070							
2140	2200	2210	2220	2230	2240		
--CCCCAAGGAGGACCGACACGCC--C---GAAGGAAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA	TT	TT	TT	TT	TT		
ATACGCCAACAGGCCGGCACTGCGAACAGAGAAGGCAAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACA	3070	3080	3090	3100	3110	3120	3130
3140							
2250	2260	2270	2280	2290	2300		
GATC----CATTCGATTAGTAAACGGATCCTTAGCACTTATCTGGGACGATCTGGAGC--CTTG---TGC	TT	TT	TT	TT	TT		
GCTCTGGCCTTGGAGATAAGAATATTCATTTCACTGATCCG---CCAACTGATAACGCCCTTTGACTTGG	3140	3150	3160	3170	3180	3190	3200
3210							
2310	2320	2330	2340	2350	2360	2370	
CTCTTCACGATACCGACCGCTTGGAGAGACTTACTCTTGATTGTA-ACGAGGATTGTGGAACTCTGGGACGCAG	TT						
CTATTCAAGCAA----CTGCAGAACCTTGCTATCGAGAGCATACCA-GATCCTCCAACCAATACTCCAGAG	3210	3220	3230	3240	3250	3260	
3270							
2380	2390	2400	2410	2420	2430	2440	
GGGTTGGGAACGCCCT-CAAA-TATGGGTGGAAATCTCCTACAGTATTGGAGTCAGA-GGAACCTAAAG	TT						
GCTCTCTCCGACCGCTACGAAGGATTGAG---AAGTCCT-CAGGACTGAACTGACCTACCTACAA	3270	3280	3290	3300	3310	3320	X

ID ERSTVMMXX standard; DNA; 3646 BP.
 XX
 AC Y00277:
 XX
 DT 17-SEP-1987 (annotation)
 XX
 DE Simian immunodeficiency virus STLV-III(MAC) proviral genome
 XX
 KW art gene; env gene; envelope protein; gag gene; genome;
 KW overlapping genes; pol gene; tat gene.
 XX
 OS Simian immunodeficiency virus
 OC Herpesviridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-8646)
 RA Chakrabarti L. ;
 RT ;
 RL Submitted (19-AUG-1987) to the EMBL Data Library.
 XX
 RN [2] (bases 1-8646)
 RA Chakrabarti L. , Guyader M. , Alizon M. , Daniel M. D. ,
 RA Desrosiers R. C. , Tiollais P. , Sonigo P. ;
 RT "Sequence of simian immunodeficiency virus from macaque and its
 RT relationship to other human and simian retroviruses";
 RL Nature 320:543-547(1986).
 XX
 CC #source: library=infected HUT78 cells; clones=lambda SIV1;
 XX
 FH Key From To Description
 FH
 FT SITE 1 175 LTR (R)
 FT RPT 1 318 5' LTR
 FT SITE 176 318 LTR (U5)
 FT SITE 488 2058 ORF (gag)
 FT CDS 551 2058 gag gene product (AA 1-506)
 FT CDS 1726 4833 pol gene product (AA 1-1056)
 FT SITE 4783 5467 ORF (R)
 FT CDS 4926 5467 R gene product (AA 1-214)
 FT SITE 5295 5633 ORF (X)
 FT CDS 5298 5633 X gene product (AA 1-112)
 FT CDS 5637 5639 R gene product (AA 1-101)
 FT SITE 5758 6083 ORF, exon 1 (tat)
 FT CDS 5788 6083 tat gene product (AA 1-99)
 FT SITE 6011 6083 (6083 is 2nd base in codon)
 FT CDS 6014 6083 ORF, exon 1 (art)
 FT SITE 6087 8735 art gene product (AA 1-23)
 FT IVS 6084 8300 (6083 is 1st base in codon)
 FT SITE 6087 8735 ORF (env)
 FT SITE 6084 8300 intron I (tat and art)
 FT CDS 6080 8735 (8301 is 3rd base in codon)
 FT SITE 8247 8551 envelope protein (AA 1-882)
 FT SITE 8251 8394 ORF, exon 2 (art)
 FT SITE 8258 8300 ORF, exon 2 (tat)
 FT CDS 8301 8384 in-frame stop codon (env)
 FT CDS 8301 8551 tat gene product (AA 100-130)
 FT CDS 8301 8551 art gene product (AA 24-108)
 FT SITE 8560 8357 (8301 is 2nd base in codon)
 FT CDS 8572 8357 ORF (F)
 FT SITE 8557 8471 F-gene product (AA 1-262)
 FT SITE 8557 8471 LTR (U3)
 FT RPT 8557 8646 3' LTR
 FT PRIM 8383 8377 enhancer-like sequence
 FT PRIM 8382 8381 enhancer-like sequence
 FT SITE 8392 8402 conserved sequence
 FT SITE 8404 8412 conserved sequence
 FT SITE 8414 8424 conserved sequence

FT SITE 9472 9646 LTR (R)
 FT PRM 9491 9498 pot. promoter sequence
 XX
 SQ Sequence 13646 BP; 3250 A; 1817 C; 2387 G; 2152 T; 0 other;
 Initial Score = 290 Optimized Score = 1240 Significance = 0.00
 Residue Identity = 55% Matches = 1422 Mismatches = 890
 Gaps = 273 Conservative Substitutions = 0

X	10	20	30	40	50	60
ATGAAAGTGAAGGAGAAATA	CAACCTTGCGAGATGGGGGTGGA	--AATGGGGCACCATGCTCCTTGGG				
TTTATGCCATCTTG-CTTCTAAGTGTCTATGG-GAT-TTATTGTATTCAATATGTCA-CAGTCT-TTTATG						
X 6120	6130	6140	6150	6160	6170	6180
70	80	90	100	110	120	130
ATATTTGATGATC TGATG---TGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAG						
GTGTACCCAGCT- TGGAGGAA TGCAGACAATTCCCCCTCTTCTGTGCAACCAAGAATAGGGATACT-TGGGGAAC						
6190	6200	6210	6220	6230	6240	6250
140	150	160	170	180	190	200
GAACCA--ACC-ACCACTCTATTTTG-TGCATCAGATGCTAAAGCAT-ATGATACAGAGGTACATAATGTTT						
AACTCAGTGCCTACCAAGATAATGATGATTTCAGA-ATTGGCCCTTAATGTTACAGAAAGCTTGTGCTT						
6260	6270	6280	6290	6300	6310	6320
210	220	230	240	250		
GG---CCCAC---ACATGC--CTGTGTACCCACAG---ACCC---CAACCCACAAAGAAGTAGTATTG						
GGGAGAAATACAGTCACAGAAACAGGCAATAGAGGACGTATGGCAACTCTTGAGACCTCAATAA--AGCCTTG						
6330	6340	6350	6360	6370	6380	6390
260	270	280	290	300	310	320
-GTAATGTCACAGAAAATTAAACA-TGTG-GAAAAATGACATGGTAGAACAGATGCATGAGGATATAATC						
TGTAAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGTGAG-ACAGATAAATGGGGAT-TGACA						
6400	6410	6420	6430	6440	6450	6460
330	340	350	360	370	380	390
AGTTTAT-GGATCAA-AGCCTAAAGCCATGTGTAATTAACCCACTCTGTGTTAGTTAAAGTGCAC						
AAATCATCACAACAAACAGCATCAA--CAACAAACAACAAACAGCAAAATCAG-TAGAGACAAGAG-AC--						
6470	6480	6490	6500	6510	6520	
400	410	420	430	440	450	460
ATTGGGGAAATCTACTAATACCAATACTAG-TAATACCAATAGTAGTAGCAGGGAA-ATGA-TGATGGAGA						
AT--AGTCATAAGACTAGT-CCTTGTGTAGTTCATGATAATTGCACAGGCTTGGAACAAAGAGCCAATGATA						
6530	6540	6550	6560	6570	6580	6590
470	480	490	500	510	520	530
AAGGAGAGATAAAAAACTGCTTTCAATATCAGCACACAAGNATAAGAGGTAAAGGTGCAGAAAG--AATATGC						
AGC1GTAAATTCAACATGACAGGTTAAAAAGAG-ACAAGAAAAAGGAGTACAATGAAACTTGGTACTCTGC						
6600	6610	6620	6630	6640	6650	6660
540	550	560	570	580	590	600
--ATT-TTTTATAAAACTTGTATATAATACCAATAGATAATGATACTACCAAGCTATACGTTGACAAGTTGTAA						
AGATTGTTGGTTGTGAAAC--AAGGGAAATAGCACT-GGTAATGAAAGTAGATGTTACA---TGAATCACTGTAA						
6670	6680	6690	6700	6710	6720	6730
610	620	630	640	650	660	670
CACCTCAGTCATTACACAGGCGCTGT-CCAAAGGTATCCTTGAGGCCAATTCCOCATACA-TTATTGTGCCCG						

TACTCTGTTATCCAAGAGTCTTGTCACAAAGAT-TATTGGGATGCTATT-AGATGTAGATATTGTGCACCT
 6740 6750 6760 6770 6780 6790 6800
 680 680 700 710 720 730 740
 GCTGTTTCCATTCTAAAATGTAATAATA-AGACGTTCAATGGAACAGGACCA--TGTACAAATGT-CAG
 :
 CCAAGTTATGCTTCTAGATGTAATGACACAAATTATTCAAGGCTTTATGCCTAAGTGTG
 6810 6820 6830 6840 6850 6860 6870
 750 760 770 780 790 800 810
 CACAGTACAATGACACATEGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAG
 :
 GTCCTTC-ATG---CACAAGGGATGATGGAGACACAG-ACT-TCTACTTGGTTCGGTTAATGGAACTAGAG
 6880 6890 6900 6910 6920 6930 6940
 820 830 840 850 860 870 880
 CAAAGA-AGAGGTTAGTAATAGATCTGCCAATTTCACAGACAATGCTAAAACCATAA-TAGTACAGCTGAA
 :
 CAGAAAAATAAAACCTATATTAA---CTGGC-ATGGTAGAGAGATAA---TAGGACTATAATTAGTCTAAAT-AA
 6950 6960 6970 6980 6990 7000
 890 900 910 920 930 940 950
 CCAAT---CTGTAGAAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGGG
 :
 GCAATATAATCTAACAAATGAAATGTAAGAAGA-CCAGGA--AAT--AAG--ACAGTTT---TA-CCA---GTC
 7010 7020 7030 7040 7050 7060
 960 970 980 990 1000 1010
 ACCA---GGGAGAGCATT---TTTACAATAGGAAAAATAGGAAATATGAGAC--AAGCA--CATTGT-AACAT
 :
 ACCATTATGTCATGGTTTCCACT--CACAACCAAGTCAATGAGAGGGCAAAGCAGGGCATGGTGTAGGT
 7070 7080 7090 7100 7110 7120 7130
 1020 1030 1040 1050 1060 1070
 TAGTAGAGGCAAAATGCAATGCCACTTTAAACAGAT---AGCTAGCAAATT---AAGAGAAACA---AT-TTG
 :
 T---TGGAGGGAAATTGGAAGGAGGCAATAAAAGAGGTGAAGC-AGACCATTGTCAAACATCCCAGGTATACTG
 7140 7150 7160 7170 7180 7190 7200
 1080 1090 1100 1110 1120 1130 1140
 GAAATAATAAAAC-AAT--AATC---TTAACGCAATCCT-CAGGAGGGGACCCAGAAATTGTAACGCACA-GT
 :
 GAACTAACAACTGATAAAATCAATTGACGGCTCCTAGAGGAGGAGATCGGAA---GTTACCTTCATGT
 7210 7220 7230 7240 7250 7260 7270
 1150 1160 1170 1180 1190 1200
 ---TTTAATTGTTGGAGGGAAATTCTACTGTAATTCAACACAACGTTAAATAGTACTTGGTTA--ATAG
 :
 GGACAAATTGCAAGAGGAGGTTCTACTGTA---AA-TGAATTGGTTCTA-AATTGGGTAGAAGAGATAG
 7280 7290 7300 7310 7320 7330
 1210 1220 1230 1240 1250 1260 1270 1280
 TACCTGGAACTACTGAAAGGGTCAAAATAACACTGAAAGGAAGTGACACAATCACACTCCATGCAGAATAAAACA
 :
 GAAGTCTAACTAACCAAGGCAAAAGGA-ACGGCATAAAAGG---AATTAC-GTACCATGTCAATTAGACA
 7340 7350 7360 7370 7380 7390 7400
 1290 1300 1310 1320 1330 1340 1350
 ATTATATAACATGTCAGGAAAGTAGGAAAAGCAATG---TATGCCCTCCCATCAGCGGACAAATTAGATGT
 :
 AATATCAACACTTGGCATAAAGTAGGCAAA---AATGTTTATTGACGGACAAAGAGAGGGAGACCTCACGTGT
 7410 7420 7430 7440 7450 7460 7470
 1360 1370 1380 1390 1400 1410 1420
 TCATCAAAATTTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACAAACATGGGTCCGAG-ATCTTCAG

CAAATCTAGCTAAGTTAAAGCAGGGGTATAAGCCAGTGTCTC-TTCCCCACCCCTCTTATTCCAGTAGAC
 8240 8250 8260 8270 8280 8290 8300
 2180 2190 2200 2210 2220 2230 2240
 CCAACCCCGAGGAGACCCGAU-AGGCC---C---GAAGGAATAGAAGAAGAAGGTGGAGAGAGACAGAGA
 ||| ||| | ||| | ||| | ||||| | + ||| ||| ||| ||| ||| ||| :
 CCATACCCGAGGATCCGCGCTCTGCCAACCAAAGAAGGGCAAAAAGGAGACGGTGGAGGCAGCGGTGGCAA
 8310 8320 8330 8340 8350 8360 8370
 2250 2260 2270 2280 2290 2300
 CAGATC---CATTCGATTAATGAAAGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGC---CTTG---T
 ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 CAGUTCCCTAGCCTTECAGATAGAATATATTCAATTTC-CTGATCCG---CCAACTGATAACGCCTCTTGACTT
 8380 8390 8400 8410 8420 8430 8440
 2310 2320 2330 2340 2350 2360 2370
 GGCCTTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTA-ACGAGGGATTGTTGGAACTTCTGGGACGC
 ||| ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 GGCTATTCAACAA-----CTGAGAACCTTGCTATCGAGAGCATACCA-GATCCTCCAACCAAATATTCCAG
 8450 8460 8470 8480 8490 8500
 2380 2390 2400 2410 2420 2430 X
 AGGGGGGTGGAAAGGCCCT-CAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAG-GGAACCTAAAG
 ||| | | | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 ABBCTCTCTGCAACCCCTACGGAGAATTGGAGAA-GTCCT-CAGGCTTGAAC TGACCTACCTACAA
 8510 8520 8530 8540 8550 8560 8570

7. KUNZ-158-CL33. SEQ

RESIVAXX Simian immunodeficiency virus STLV-III(AGM) provir

ID RESIVAXX standard; DNA; 9264 BP.
 XX
 AC Y00295;
 XX
 DT 28-SEP-1987 (annotation)
 XX
 DE Simian immunodeficiency virus STLV-III(AGM) proviral genome
 XX
 KW env gene; envelope glycoprotein; gag gene; genome;
 KW long terminal repeat; overlapping genes; pol gene; protease;
 KW R gene; reverse transcriptase; sor gene; transfer RNA;
 KW transfer RNA-Lys; unidentified reading frame.
 XX
 OS Simian immunodeficiency virus
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-9264)
 RA Reitz M.;
 RT ;
 RL Submitted (17-JUL-1987) on tape to the EMBL Data Library by:
 RL Reitz M., LTCB, NIH, Bldg 37 Room 6C09, Bethesda MD 20892, USA.
 XX
 RN [2] (bases 1-9264)
 RA Franchini G., Gallo R.C., Guo H.G., Gurgo C., Callatti E.,
 RA Fagnoli K., Hall L., Wong-staal F., Reitz M.S.;
 RT "Sequence of simian immunodeficiency virus and its relationship to
 RT the human immunodeficiency viruses";
 RL Nature 328:539-543(1987).
 XX
 CC *source: library=EMBL-3; cell line=infected K6W
 XX
 FH Key From To Description
 FH
 FT SITE 1 176 R-region of 5'-LTR
 FT 5'LTR 1 302 5'-long terminal repeat

FT	SITE	152	157	put. polyA signal
FT	SITE	177	303	U3-region of LTR
FT	tRNA	306	323	transfer RNA-Lys(3)
FT	CDS	559	883	protein p17 (AA 1 - 115)
FT	CDS	589	2057	gag gene product
FT	CDS	884	2056	protein p24 (AA 1 - 391)
FT	CDS	1714	2263	protease (AA 1 - 184)
FT	CDS	1714	4875	pol gene precursor polypeptide
FT	RPT	1725	1756	imp. direct repeat 1
FT	RPT	1788	1819	imp. direct repeat 1
FT	CDS	2266	4875	reverse transcriptase (AA 1 - 870)
FT	SITE	4552	4656	polypurine tract
FT	CDS	4778	5448	sor gene product
FT	CDS	5702	7305	large envelope glycoprotein gp120 (AA 1 - 536)
FT	CDS	5702	8371	env gene product
FT	CDS	7310	8371	small envelope glycoprotein gp32 (AA 1 - 354)
FT	CDS	8196	8828	3'-ORF (AA 1 - 211)
FT	SITE	8376	8580	polypurine tract
FT	SITE	8581	9020	U3-region of LTR
FT	RPT	8581	9264	3'-long terminal repeat (LTR)
FT	SITE	9020	9264	R-region of 3'LTR
FT	PRM	9059	9063	put. TATA-box

XX
SQ Sequence 5264 BP: 3121 A; 1749 C; 2309 G; 2085 T; 0 other;

Initial Score = 294 Optimized Score = 1240 Significance = 0.00
 Residue Identity = 54% Matches = 1421 Mismatches = 897
 Gaps = 266 Conservative Substitutions = 0

X	10	20	30	40	50	60	
ATGAGAGTGAACGGAGAATAATCAGCACTT--GTGGAGATGGG---GGTGGAA---AATGGGGCACCATGTC							
ATCAGCTGCTTATCGCCATCT-TGCTTTAAAGTGCTATGGGATCTATTGTACTCAATATGTCA-CAGTCT-	X 5750	5760	5770	5780	5790	5800	5810

```

        140      150      160      170      180      190
TGGAAAGGAAGCA---ACC-ACCACTCTA1TTTG-TGCATCAGATGCTAAAGCAT-ATGATAACAGAGGTACATA
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGGAAACAACACTCACTGCGCTAACGAGATAATGGTGATTATTCAAGA-ATTGGGCCCTTAATGTTACAGAAAAGCTTTG
      5890      5890      5890      5890      5890      5890      5890      5890

```

280 280 300 310 320 330
 -GAAATTAAACATGTGGAAAA--ATGA-CAT--GGTAGAAC---AGATGCATGAGGGATATAATCAG-
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 CCTCAATAAAAGCTTGTGTAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATA
 6020 6030 6040 6050 6060 6070 6080

340 350 360 370 380 390
 TAT-GGGAT---CA-AAGCTAAA-GCCATGTGTAAGATTAAACCCC-AC-TCTGTGTTAGT-TTAAAGTGCA
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GAT-GGGATTCACAAAAATCATCAACAAATAACAAACAGCAGCACCAACATCAGCACCAGTATCAGAAAAAA
 6090 6100 6110 6120 6130 6140 6150

400 410 420 430 440 450 460
 CTCATTTGGGAACTACTAATAACCAATCTAG-TAATACCAATAGTAGTAGCGGGGAA-ATGA-TGATGG
 ||| |||| |||| ||| ||| |||| ||| |||| ||| |||| ||| ||| |||
 TAGACATGCTAATGAGACTAGTTCTTGTA-TAGCTCAGAATAATTGCACAGGCTTGGAACAGAGCAAATG
 6160 6170 6180 6190 6200 6210 6220

 470 480 490 500 510 520 530
 AGAAAAGGAAAGA-TAAANAAACTCTCTCTTCAATATCAGCACAGNATAAGAGGTAAAGGTGCAGAAAG--AATA
 ||| |||| ||| ||| ||| |||| ||| |||| ||| ||| ||| |||
 ATAAGCTGTAAATTCAACCAGACAGGGTTAAAAGAG-ACAAGACAAAGGAGTACAATGAAACTTGGTACTC
 6230 6240 6250 6260 6270 6280 6290 6300

 540 550 560 570 580 590 600
 TGC---ATTT-TTTTATAAAACTTGATATAATACCAATAGATAATGATACTACCAAGCTATACTGTTGACAAGTTG
 ||| |||| ||| ||| ||| |||| ||| |||| ||| ||| |||
 TACAGATTGGTTTGTAAC--AAGGGAATAGCACT-GATAATGAAAGCAGATGCTACA---TGAATCACTG
 6310 6320 6330 6340 6350 6360

 610 620 630 640 650 660 670
 TAACACCTCACTCATTACACAGGCTGT-CCAAAGGTATCCTTGAGCCAATTCCCACATTATTGTGC
 ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TAACACTTCTTTATCCAAAGAGTCCTGTGACAAACAT-TATTGGGATACTATTAGATTAGGTATTGTGCAC
 6370 6380 6390 6400 6410 6420 6430

 680 690 700 710 720 730 740
 CGGCTGGTTTGCGATTCTAAATGTAATAATA-AGACGTTCAATGGAACAGGAC--CATGTACAAATGT-C
 ||| |||| ||| ||| ||| ||| ||| ||| ||| |||
 CTCCAGGTATGCTTGTAGATGTAATGACACAAATTATTCAAGGCTTATGCCTAAATGTTCTAAGGTGG
 6440 6450 6460 6470 6480 6490 6500

 750 760 770 780 790 800
 ABCACAGTACAAATGACACATGGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TGGCTCTTTC-ATG--CACACGGATGATGGAGACACAG-ACT-TCTACTTGGTTGGCTTAAATGGAACTAG
 6510 6520 6530 6540 6550 6560 6570

 810 820 830 840 850 860 870
 AGCAGAAGA-AGAGGTAGTAAATTAGATCTGCCAATTTCACAGACAATGCTAAAACCATAA-TAG--TACAGC
 ||| |||| ||| ||| ||| ||| ||| ||| ||| |||
 AGCAGAAAATACAACTTATATTAA---CTGGC--ATGGTAGGGATAA---TAGGACTATAATTAGTTAAATA
 6580 6590 6600 6610 6620 6630 6640

 880 890 900 910 920 930 940
 TGAACCAATCTGAGAAATTAAATTGATACAAGACCCAAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 AGTATTATAATCTAACAAATGAGAAGA-CCAGGA--AAT--AAG--ACAGTTT---TA-CCA---GT
 6650 6660 6670 6680 6690

 950 960 970 980 990 1000 1010
 GACCA--GGGAGAGCAATT--TGTACAAAT-AGGAAAAATAGGAAATATGAGACAAGCA--CATTGTAACATT
 ||| ||| ||| ||| ||| ||| ||| |||
 CACCAATTATCTGGATTCTTCCACTCACACCCACTCACTGATAAGGCCA-AAGCAGGCATGGT---GTT
 6700 6710 6720 6730 6740 6750 6760

 1020 1030 1040 1050 1060 1070
 AG---TAGACCAAAATGCAATGCCACTTTAAACAGAT-AGCTAGCAAATT---AAGAGAACAA-AT-TTG
 ||| |||| ||| ||| ||| ||| ||| |||
 GGTTTGGAGGAAATGAGATGCAATAAAAGAGGTGAAACAGACCAATTGTCAAACATCCCAGGTATACTG
 6770 6780 6790 6800 6810 6820 6830

 1080 1090 1100 1110 1120 1130 1140
 GAAATAATAAAC-AAT--AATC--TTAACCAATCCT-CAGGAGGGGACCCAGAAATTGTAACGCACA-GT
 ||| |||| ||| ||| ||| ||| ||| |||
 GAACTAACATGATGAAATAATCAATTAAACGCTCCTGGAGGAGGAGATCCGGAA---GTTACCTTCATGT
 6840 6850 6860 6870 6880 6890 6900

1150 1160 1170 1180 1190 1200
 ---TTTAATTTGAGGGGAATTTCCTACTGTAATTCAACACAACGTGTTAATAGTACTTGTTA--ATAG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GGACAAATTCGAGGAGAGTTCCTCTACTGTA---AA-TGAATTGGTTCTA-AATTGGGTAGAGGATAG
 6910 6920 6930 6940 6950 6960 6970
 1210 1220 1230 1240 1250 1260 1270 1280
 TACTTGGAACTACTCTGAGGGTCATAAACACTGAGGAAGTGACACAATCACACTCCCATGCAGAATAAAA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GGATGTAATACCCAGAGGCCAAGA-AC---GUCATAGAAGGAATTAC-GTGCCTGTCATATTAGACA
 6980 6990 7000 7010 7020 7030 7040
 1290 1300 1310 1320 1330 1340 1350
 ATTATATAAAGATTTGCGAGAACTGAAAGAAATG---TATGCCCTCCCATCAGCGGACAAATTAGATGT
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 AATAATCAGACATTGCGATAAAGTAGGCAA---AATGTTATTGCGCTCCAAGAGAGGGAGACCTCACGTGT
 7050 7060 7070 7080 7090 7100 7110
 1360 1370 1380 1390 1400 1410 1420
 TCATCAAATATTACAGAGCTCTATTAAACAAGAGATGGTGGTAATAACAAACAATGGGTCCGAG-ATCTTCAG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 AACTCCACAGTGGACAGTCTCGATAAGCATAGATTGGACTGATGGAAACCA---AACTAGTATC---
 7120 7130 7140 7150 7160 7170
 1430 1440 1450 1460 1470 1480 1490
 ACCCTGGAGGAGGAGATATGAGGGACAAATTGGA---GAAGTGAATTATAAAATATAAGTAGTAAAAATTGAA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 ACCATGAA-ATGCGAG---AGGTGGCGAGAACTGTATCGATTAGAGTTGGGAGATTATAAATTAGTAGAGATCACT
 7180 7190 7200 7210 7220 7230 7240
 1500 1510 1520 1530 1540 1550
 CCATTAGGCTTACACCCACAGGAGGAG---AAGAGTGGT-GCA---GAGAGAAAAAGAGCAGTGGG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 CCGAGTGGGCTTGGCGGGGACAGATGAGTGGAGGGTACACTACTGGTGGCACCTCAAGAAATAAAAG-AG-GGG
 7250 7260 7270 7280 7290 7300 7310
 1560 1570 1580 1590 1600 1610 1620
 AATAGGAGGTTTGTCTTGGTTCT-TGGGAAGCAGCAGGAAGCACTATGGCGCACGGTCAATGACGCTGA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 TCTTGTGCTADGGTCTTGAGTTCTCGCAACGGCAGGTTCTGCAATGGCGCGTGTGTTAGGCTGA
 7320 7330 7340 7350 7360 7370 7380
 1630 1640 1650 1660 1670 1680 1690
 CGGTACAGAGGAGGAGATTATTGCTGGTATAGTGCAGCAGCAGAACATTGCTGAGGGCTATTGAGGGCGC
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 CCGCTCAAGCAGGAGACTTTATTGCTGGGAGATAGTGCAGCAACAGCAACAGCTGTTGGCGTGGTCAAGAGAC
 7390 7400 7410 7420 7430 7440 7450
 1700 1710 1720 1730 1740 1750 1760 1770
 AACAGAGATCTTGCAGACTCACAGTCTGGGAGATCAAGCAGCAGAACATTGCTGAGGGCTATTGAGGGCGC
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 AACAGAGAAATGATGCGATTCACCGTCTGGGAGAACAAAGAACCTCAGACTAGGGTCACTGCCATCGAGAAAGT
 7460 7470 7480 7490 7500 7510 7520
 1780 1790 1800 1810 1820 1830 1840
 ACCCTAAAGGAGTCAAGCTCTGGGAGATTGGGGTTGCTCTGGAAAACCTCATTGCAACACTGCTGTGCCTT
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 ACTTACAGAGGAGGAGGAGCTGAAATGCTTGGGGAGTGCGTTAGACAAGTCTGCCACACTACTGTACCAT
 7530 7540 7550 7560 7570 7580 7590 7600
 1850 1860 1870 1880 1890 1900 1910
 GG---AATGCTTGTGGAGTAATAATCTCTGGAACAGATTTGGAAATAACATGACCTGGATGGAGTGGGACA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GCGCAAAATGCGAT---CTACGACAGAGACTGGAACA-AT---GA-TA-----CTGGCAAGAGTGGGAGC
 7610 7620 7630 7640 7650 7660 7670 7680

1920 1930 1940 1950 1960 1970 1980
 GAGAAATTGAC-AATTACACAAGCTTAATACATTCCTTAATTGAAGAATCGCAAAACCGAGCAAGAAAAGAAT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GAAGAAGTTRAC1TCCTGAGGAAATATAACA-GCCCTCCTAGAAGAGGGCACAAATTCAACAAAGAGAAGAAC
 7660 7670 7680 7690 7700 7710 7720

 1990 2000 2010 2020 2030 2040 2050
 GAACAGAAATTATTGGATTAGATAAAATGGGCAAGTTGTGG-AATTGGTTAACATAACAAATTGGCTGTG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ATGTATGAAATTACAAAAAGTTGAATAGCTGGG-ATGTGTTGGCAATTGGTTGACCTTGCTCTGGATAAA
 7730 7740 7750 7760 7770 7780 7790 7800

 2060 2070 2080 2090 2100 2110 2120
 GTATATAAAATATTCA--TAATGATAGTAGGAGGGCTTGGTAGGTTAACAGATAGTTTGCTGTACTTTCT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GTATAT-AGAAATATGAAATTATC-TAGTTGTAGGAGTAATACTGTTAACAGATAGTGATCTATATAGTACAA
 7810 7820 7830 7840 7850 7860 7870

 2130 2140 2150 2160 2170 2180
 ATAGTGAATAGAGTT-AGCCAGGGATATTCAACCA---TTATCGTTTCAGACCCAC---CTCCCA---AC--
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ATGCCTAGCTA-AGTTAAGGCAGGGTATAGGCCAGTGTTC-TTCCCCCACCTCTTATTCCAGTAGACTC
 7880 7890 7900 7910 7920 7930 7940

 2190 2200 2210 2220 2230 2240
 ---CCCGAGGGACCCGACA-CGCC---C---GAAGGAATAGAACAGAACAGGTGGAGAGAGACAGAGACA
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ATACCCAAACAGGACCCGGCACTGCCAACCCAGAACAGAACAGGAAAGGAGACGGTGGAGAACGGCGGTGGCAACA
 7950 7960 7970 7980 7990 8000 8010

 2250 2260 2270 2280 2290 2300
 GATC----CATTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGC--CTTG---TGC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GCTCCTGGCCTTGGCAGATAAGAATATAATCATTTC-CTGATCCG---CCAACTGATAACGCCCTTGTACTGG
 8020 8030 8040 8050 8060 8070 8080

 2310 2320 2330 2340 2350 2360 2370
 CTCTTCAGCTAACACCCCTTGAGAGAC1TACTCTTGATTGTA-ACGAGGATTGTGGAACTTCTGGGACGCAG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTATTCAGGAA-----CTGCAGAACCTTGCTATCGAGAGGATACCA-GATCCTCCAACCAATACTCCAGAG
 8090 8100 8110 8120 8130 8140

 2380 2390 2400 2410 2420 2430 2440
 GGGGATGGGAAAGGCGT-CAAA-TATGGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACCTAAAG
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 GCTCTCTGCGAACGCTACGGAAAGGATTCGAG--AAGTCCT-CAGGACTGAACCTACCTACAA
 8150 8160 8170 8180 8190 8200 X

8. KUNZ-158-CL33. SEQ

HIV2R0DX Human immunodeficiency virus type 2 ROD isolate RN

ID HIV2R0DX standard; RNA; 5671 BP.
 XX
 AC X05291:
 XX
 DT 04-JUN-1997 (annotation)
 XX
 DE Human immunodeficiency virus type 2 ROD isolate RNA genome
 DE (HIV-2)
 XX
 KW acquired immune deficiency syndrome; art gene; env gene; f gene;
 KW gag gene; pol gene; q gene; r gene; tat gene.
 XX
 OS Human immunodeficiency virus type 2

OS ROD isolate
 OC Viridae; ds-RNA enveloped viruses; Retroviridae.
 XX
 RN 111 (bases 1-8671)
 RA Alison M.
 RT ;
 RL Submitted (03-JUN-1987) on tape to the EMBL Data Library by:
 RL Marc Alison, Unité d'Oncologie virale, and CNRS UA1157, Institut
 Pasteur, 25 rue du Dr Roux, 75724 Paris CEDEX 15, France.
 XX
 RN 121 (bases 1-8671)
 RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,
 RA Alison M.;
 RT "Genome organization and transactivation of the human
 RT immunodeficiency virus type 2";
 RL Nature 326:652-655(1987).

XX
 RN 131
 RA Clavel F., Guyader M., Guetard D., Salle M., Montagnier L.,
 RA Alison M.;
 RT "Molecular cloning and polymorphism of the human immune deficiency
 RT virus type 2";
 RL Nature 324:681-685(1987).

FH	Key	From	To	Description
FT	SITE	1	173	R region
FT	RPT	1	253	LTR
FT	SITE	1	8671	HIV-2 RNA corresponding to integrated proviral DNA
FT	SITE	174	299	U5 region
FT	SITE	303	320	primer binding site
FT	CDS	546	2111	gag protein
FT	CDS	1829	4935	pol protein
FT	SITE	4513	4626	polypurine tract 2
FT	CDS	4629	5513	q protein
FT	CDS	5582	5896	r protein
FT	CDS	5843	6140	tat protein part 1 (6140 is 2nd base in codon)
FT	CDS	6071	6140	art protein part 1 (6140 is 1st base in codon)
FT	CDS	6147	8720	env protein
FT	CDS	8307	8400	tat protein part 2 (8307 is 3rd base in codon)
FT	CDS	8307	8536	art protein part 2 (8307 is 2nd base in codon)
FT	CDS	8557	9324	f protein
FT	SITE	8925	8935	polypurine tract 1
FT	SITE	8942	8487	U3 region
FT	RPT	8342	9871	LTR
FT	PRM	9329	9339	core enhancer sequence
FT	PRM	9401	9416	core enhancer sequence
FT	SITE	9420	9427	pot. SP1 factor binding site
FT	SITE	9428	9437	pot. SP1 factor binding site
FT	SITE	9438	9443	pot. SP1 factor binding site
FT	PRM	9465	9470	TATA-box
FT	SITE	9483	9671	R region
FT	SITE	9543	9654	pot. polyA signal
FT	POLYA	9671	9671	polyA site

XX
 SQ Sequence 8571 BP; 3314 A; 1973 C; 2401 G; 1983 T; 0 other;

Initial Score = 306 Optimized Score = 1185 Significance = 0.00
 Residue Identity = 53% Matches = 1395 Mismatches = 888
 Gaps = 323 Conservative Substitutions = 0

X 10 20 30 40 50 60
 ATGAA---AGTGTGGAGAATTAACAGG---ACTTGCTGGAGATGGGGGGTGGAAATGGGGCACCATGCTCCTTGG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ACCAGAACAGTGTGGATGAACTGCTTATTGCCATTATTATT---AGCTAGTGC---TTG---CTTAG
 X 6140 6150 6160 6170 6180 6190
 70 80 90 100 110 120 130
 GATATTG-----ATG-ATCTGAGTGCTACAGAAAAATTGTG-GGTCACAGTCTATTATGGGGTACCTG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TATATTGCAACCCNATAATGTAACCTG-TTTCTATGGCGTACCCACGTGGAAAAATGCAACCATTCCCCTCTTT
 6200 6210 6220 6230 6240 6250 6260
 140 150 160 170 180 190
 TGTEGAAGGAGCAACCAACCACTCTATTTTG---CAT-CAG---ATGCTAAAGCAT-ATGATACAGAGGT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TGTCAA-CCAGAAATAGGGAA---TACTTGGGAAACCATACAGTGCTTGCTGACAATGATGATTATCAGG-
 6270 6280 6290 6300 6310 6320 6330
 200 210 220 230 240 250 260
 ACATAAATGTTG---GGCCACACATGCCCTGTGACCCACAGACCCCCAAC-CCACAAAGAAGTAGTATTGGTAAA
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AAATATA-CTTGTAAATGAAACAGAGCTTTG-ATGCATGGAATAATACAGTAACAGAACAGCAATAGAAGA
 6340 6350 6360 6370 6380 6390 6400
 270 280 290 300 310 320 330
 TG---TGACAGAAAATTITAAUCATGTGGAAAAATGACATG-GTAGAACAGATGCATGAGGGATATAATCAGTT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 TGCTGGCATCTATTGAGACAT---CAATAAAACCATGTGTCAAACTAACACCTTATGTGT-AGCAATGA
 6410 6420 6430 6440 6450 6460 6470
 340 350 360 370 380 390 400
 ATGGGATCAAAACCTAAAGC---CATGTGTAAAATTAAACCCACTCTGTGTTAGTTAAAGTGCACTGATTG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AATACAGCAGCACAGAGAGCAGCACAGGGAA---CAACACAAAC-CTCAAAGAGCACAA---GCACAAACCACA
 6480 6490 6500 6510 6520 6530
 410 420 430 440 450 460
 GGGAAATGCTAAATACCAATA-CTAG---TAA-TACCAATAGTAGTAGCG---GGGAAATGA---TGATGGA
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ACCACACCCAC---AGACCACAGCAAGAGATAAGTGAGGGATACTCCATGCGCACGCGCAGACAACTGCT-CA
 6540 6550 6560 6570 6580 6590 6600
 470 480 490 500 510 520 530
 GAAAAGGAGAGATAAAAACGTGCTTTCAATATCAGCACAGNATAAGAGGG-TAAGGTGCAGA-AAGAATAT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 GGATGGGAGAGGAAACAAACATC---AATTGCCAGTTCAA-TATGACAGGATTAGAAAGAGATAAGAAAAAA
 6610 6620 6630 6640 6650 6660 6670
 540 550 560 570 580 590 600
 GCAATTTTTATAAACTGTATATAATACCAATAAGATAATGATACTACCACTACGCTACGTTGACAAGTTGTAAC
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ACAGTATAAT-GAAACATGGTA---CTCAAAAGATGTGGTTGTGAGACAAATAATAGCACAA---ATCAG
 6680 6690 6700 6710 6720 6730
 610 620 630 640 650 660 670
 AGCTCAGTCATTACACAGGGCTCTTCCAAAGGTATCCTTGAGCCAATTCCCACATACATTATTGTGCCCCGGCT
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AGC-CAGT-CTTACATGAAACCATTGC-AACACATC---AGTCA---TCACAGA-ATCA-TGTGACAAG---C
 6740 6750 6760 6770 6780 6790
 680 690 700 710 720 730 740
 GGTTTTCGCGATTCTAAATGTTATAATA-AGACGTTCAATGGAACAGGGACC-ATGTACA---AATGTCAGCA
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 ACTATTGGGATGCTATAAGGTTAGATACTGTGCAACCACCGGGTTATGCCCTATTAAGATGTAATGATACCA
 6860 6870 6880 6890 6900 6910 6920

750 760 770 780 790 800
CACTA--CA---ATGTTACACA--TGGAAATTAGGCCAGTAGTATCAAC-TCAAC--TGCTG-TTGAATGGCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
-ATTATTCAAGGCTTTCACCCAACTGTTCTAAAG-TAGTAGCTTACATGCACCAGGATGATGGAAACGCA
6970 6980 6990 6990 6990 6990

810 820 830 840 850 860 870
GTCTAGCAGAAGAAAGAGGGTAGTAATTAGATCTGCCAATTCACAGACAATGCTAAAACCA--TAATAGTACA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AACTTCCACATG---GTTTUGCTTTA-A--TGGCACTAGAGCAGAGAATAAACATATCTATTGGCATG
6940 6950 6960 6970 6980 6990 7000

880 890 900 910 920 930 940
GCTGAACCAATCTGTAAGAANTTAATTGTACAAGACCCAAACAACAATACAAGAA--AAAGTATCCGTATCCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCAWAGATAA---TAISAACT--ATCAT-CA-GCTTAAACA--AATATTATAATCTCAGTTGCATTGTAAG
7010 7020 7030 7040 7050 7060

950 960 970 980 990 1000 1010
AGGAAACCAAGGAA--GAGCATTTCATTAC-AATAGGAAAAAT---AGGAAATATGAGACAAGCACATTGTAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACG---CCAGGAAATAAGACAGTGAACAAATAATGCTTATGTCAGGACATGTGTTCACTCCACT---AC
7070 7080 7090 7100 7110 7120 7130

1020 1030 1040 1050 1060 1070
ATTOGTAGAGCAAAATGCAATGCCACTTTAAACA---GATAG--CTA--GCAAATTAAGAGA-ACAAT-T
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
--CAGCCGATCAATA---AAAGACCCAGACAAGCATGGTGCTGGTTCAAAGGAAATGGAAAGACGCCATGC
7140 7150 7160 7170 7180 7190

1080 1090 1100 1110 1120 1130 1140
TGAAAAT-AATAAAACAATAATCTTTAAGCAATCCTCAGG---AGGGGACCCAGAAATTGTAACGCACAGTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ABGAAGTGAAGGAAACCT---TGCAAAACATCC-CAGGTATAGAGGAACC---AAT-GACACAAGGAATA
7200 7210 7220 7230 7240 7250 7260

1150 1160 1170 1180 1190
TTAATTGTCGAG---GGGAATTTTTC-TAC---TGTA--ATTCA---ACAC-AACTGTTAATAGTAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTAGCTTTAACCGGCCAGGAAAGCTCAGACCCAGAAGTAGCATACATGTGGACTAACTGCAGAGGGAG-AG
7270 7280 7290 7300 7310 7320 7330

1200 1210 1220 1230 1240 1250 1260
TTGTTTAATAGTACTTGGAGTACTGAAAGGGTCAAAT-AACACTGAAGGAAGTGACA---CAATCACACTCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTTCTCTAUTCACAT-GACT---TGGTTCTCAATTGGATAGAGAA-TAAGACACACCGCAATTATGC-AC
7340 7350 7360 7370 7380 7390

1270 1280 1290 1300 1310 1320 1330
CATGCAGAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATG--TATGCCCTCCCATCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CGTACCATATAAGCAAAATAATTAAACACATGGCATAAGGTAGG---GAGAAATGTATATTGCTCCCA----
7400 7410 7420 7430 7440 7450 7460

1340 1350 1360 1370 1380 1390 1400
CGGACAAAATAGATGTTCTCAAATATTACAG---GGCTGC-TATTAACAAGAGATGG--TGGTAATAACAAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGA-AGGAAAGTCTCCTCAACT-CAACAGTAACCAGUATAATTGCTA-ACATTGACTGGCAA-AACAAT
7470 7480 7490 7500 7510 7520 7530

1410 1420 1430 1440 1450 1460
AATGGTCGAAAT---CTTCAGACCTGGAGGGAGG-AGA--TATGAGGGACA-ATTGGAGAAGTGAATTATA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATCAGACAAAGCTTACCTTGTG-AGAGGTGGCAGAACTAT----ACAGATTGGAGTTGGGAGAT-TA
7540 7550 7560 7570 7580 7590

1470 1480 1490 1500 1510 1520 1530
 TAAATATAAATAGTAAAAATTGACCATAGGAGTAGCACCCACCAAGGCAAAGAGA-AGAGTGTTGCAGA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 TAGAT----TGTAGAAATAACACCAATTGGCTTCGACCTACAAAAGAAAAAGATACTCCTCTGCTCA
 7600 7610 7620 7630 7640 7650 7660
 1540 1550 1560 1570 1580 1590 1600
 GAGAAAAAGAGCAGTGG-GAAATAGGAGCTTGTTCCTGGGTTCT-TGGGAGCAGCAGGAAGCACTATGGG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 CGGAGACATACAGAGGGTGTTGTTGACTAGGGTTCTGGGTTCTCGCAACAGCAGGTTCTGCAATGGG
 7670 7680 7690 7700 7710 7720 7730
 1610 1620 1630 1640 1650 1660 1670
 CCGACGGTCAATGACCTGACGGTACAGGCCAGACAATTATTGCTCTGGTATAAGTGCAGCAGCAGAACAAATT
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 CGGAGCGTCCCTGACCGTGTGGCTCAGTCCCGGACTTTACTGGCCGGATAGTGCAGCAACAGCAACAGCT
 7740 7750 7760 7770 7780 7790 7800
 1680 1690 1700 1710 1720 1730 1740 1750
 GCTCAAGGCGTAAAGAGCAACAGCAGCATCTATTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GTTGGACGTTGCAAGAGACAACAGAACTGTTGCGACTGACCGTCTGGGGACGAAAAACCTCCAGGCAAG
 7810 7820 7830 7840 7850 7860 7870
 1760 1770 1780 1790 1800 1810 1820
 AATCCTGGCTGTGGAAAGATAACCTAAAGGATCAACAGCCTCTGGGATT--TGGGGTTGCTCTGGAAAACTC
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 ATGCACTCTATAGAAAGTAACCTACAGACCCAGGGGGCT--AAATTCACTGGGATGTGCGTTAGACAA
 7880 7890 7900 7910 7920 7930 7940
 1830 1840 1850 1860 1870 1880 1890
 ATTTCACGCAATTGCTGTGCGTGGAAATGCTAGTTGGAGTAATAAAATCTCTGGAACAGATTTGGAATAACATG
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GTCTGCCACACTACTGTACCATGG-----GT-----TAATGATTCCCTAGCACCTGACTGGGACAATATG
 7950 7960 7970 7980 7990 8000
 1900 1910 1920 1930 1940 1950 1960
 ACCTGGGATAGACTGGGACAGAGAAATTAAACAATTACACAAGCTTAATA-CATTCCCTTAATT-GAAGAACATGC
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 AGCTGGCAAGATGGGAAAAACAAAGT--CCGCTACCTGGGAGGCAAATATCAGTAAAAGTTAGAACAGGGCAC
 8010 8020 8030 8040 8050 8060 8070
 1970 1980 1990 2000 2010 2020 2030
 AAAAUCCAGCAAGAAAGAAATGAAACAAGAATTATTGGAATTAGATAAAATGGCAAGTTGTGGAATTGGTTTA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 AAGATTCAGCAAGAAGAAAAATATGATGACTACAAAAATTAAATAGCTGGGATATTTTGGCAATTGGTTTG
 8080 8090 8100 8110 8120 8130 8140 8150
 2040 2050 2060 2070 2080 2090 2100
 ACACTAACAAATTGGCTGGGTATA-TAAAAAT--ATTCTATAATGATAGTAGGAGGGCTGGTAGGTTAACAA
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 ACTTAAACCTCTGGCTGAGTATATTCAATATGGAGTGCTTATAATAGTAGCAG---TAATAGCTTTAACAA
 8160 8170 8180 8190 8200 8210 8220
 2110 2120 2130 2140 2150 2160
 TAGATTTGGCTGCTATAGTAAAGTAGAGTTAGGGCAGGGATATTCAAC-ATTATCGTTTC-----
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 TAGTGATAATGATGACTACAAATGTTAAAGTAGGCTTAGAAAGGGCTATAGGCCTGTTTCTCTCCCCCCCCCG
 8230 8240 8250 8260 8270 8280 8290
 2170 2180 2190 2200 2210 2220
 ---AGACGCAAGU-----TCCACGCCCGAGGGACC--CGACAG---GCC--CGAAGGAAATAGAAGAAGAAGGT
 ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
 GTTATATGCAACAGTCCATATCCACA-AGAACCGGGGACAGCCAGGCCAACGAAGAAACAGAACAGAACGGT
 8300 8310 8320 8330 8340 8350 8360

2230 2240 2250 2260 2270 2280 2290
 GGAGAGAGAGAGAGAGACAGATCCATTGATTAGTGAACGGATCCTTACCTTATCTGGGACGATCTGC
 ||| +
 GGAG
 8370 8380 8390 8400 8410 8420 8430
 2300 2310 2320 2330 2340 2350 2360
 AGCGTTGTTACCTCTT---CA---CTA-CCA-CGCGCTTGAGAGACTTACTCTTGATTGTAAC--GAGGATTGTG
 +
 ~---TT-CGCGCTCTTCACCGAGACTTACAGCATCTGCAGGGACTTACTATCCA--GGAGCTTCCTGACCC
 8440 8450 8460 8470 8480 8490
 2370 2380 2390 2400 2410 2420 2430
 GAACT---TCTGAGACGAGGGGGTGGGAAGCCCTCAAATA-TTGGGTGGAATC-TGCTACAGTATTGGAGTC
 ||| +
 CAGCTCATCT---AC-UAGAAATCTCAGAGACTGGCTGAGACTTAGAACAGCCTTCTTGCAATATGGGTGCGA
 8500 8510 8520 8530 8540 8550 8560
 X
 --GAACTA--AAG
 + + + + + + +
 GTGAGATCGAGAG
 8570 8580

9. KUNZ-156-CL33 SIV

RESIV251 Simian Immunodeficiency virus (Mac251) envelope gene

ID RESIV251 standard; DNA; 1142 BP.
 XX
 AC X06879; Y00254;
 XX
 DT 23-JUN-1988 (acc# added)
 DT 26-MAY-1988 (annotation)
 XX
 DE Simian Immunodeficiency virus (Mac251) envelope gene DNA (part.)
 DE integrated copy
 XX
 KW env gene; envelope gene.
 XX
 OS Simian immunodeficiency virus
 OC Viridiviridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN 111 (bases 1-1142)
 RA Kestler H. W.;
 RT ;
 RL Submitted (26-FEB-1988) to the EMBL Data Library by:
 RL Kestler H. W., Harvard Medical School, New England Regional Primate
 RL Research Center, Department of Microbiology, One Pine Hill Drive,
 RL Southborough, Mass. 01772, USA.
 XX
 RN 111
 RA Kestler H. W., Li Y., Naidu Y. M., Butler C. V., Ochs M. F.,
 RA Jaxungal D., King N. W., Daniel M. D., Desrosiers R. C.;
 RT "Comparison of Simian immunodeficiency virus isolates";
 RL Nature 331:618-622(1988).
 XX
 CC Resource: strain=Macaque muleatta 251(host); clone=Lambda SIV 251;
 XX
 FH Key From To Description
 FH
 FT SITE 1 1142 put. env gene
 (1 is 2nd base in codon)
 FT SITE 1026 1028 in frame stop codon
 XX

10. KUNZ-156-OLB4. SET

Figure 1. Structure of the env gene of HTLV-III, C

ID M15127 unannotated; xxx; 306 BP.

xx

AC

XX

DT

2

DE

XX

KW

XX

OS

□□

xx

RN

RA

RT

RT

RL

xx

FH

113 (Cases 1-306)

RA Gold W. G., Bogroski J. G., Rosen C. A., Haseltine W. A.

RT "Expression of the env Gene Protein of Human T-Lymphotropic Virus

BT Type III (HTLV-III/LAV) in Bacteria";

见于 Virology 31:633-637(1987)。

RE: [View Log](#) [Edit Log](#) [Get Log](#)

FH

FH
XX

50 Sequence 308 bp; 35 A; 68 C; 87 G; 66 T; 0 other;

Initial Score = 194 Optimized Score = 298 Significance = 0.00
Residue Identity = 89% Matches = 298 Mismatches = 2
Gaps = 1 Conservative Substitutions = 0

2140 2150 2160 2170 2180 2190 2200 2210
TACGCCAGGAAATTTCAACCATTATCGTTCAAGACCCACCTCCCAACCCCGAGGGGGACCCGACAGGCCCGAAGG
|||||
TACGCCAGGAAATTTCAACCATTATCGTTCAAGACCCACCTCCCAATCCCGAGGGGGACCCGACAGGCCCGAAGG
X 10 20 30 40 50 60 70

2220 2230 2240 2250 2260 2270 2280
AATAGAAGAAAGAAGATGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTAGCACTTAT
|||||
AATAGAAGAAAGAAGGTGAGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTAGCACTTAT
80 90 100 110 120 130 140

2290 2300 2310 2320 2330 2340 2350
CTGGGGACGATCTTGCGGAGCCCTTGTGCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGA
|||||
CTGGGGACGATCTTGCGGAGCC-TGTGCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGA
150 160 170 180 190 200 210

2360 2370 2380 2390 2400 2410 2420
GGATTGTGGAACCTCTGGGACCCAGGGGGTGGGAAGGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGA
|||||
GGATTTGTGUAACCTTC1GGGAUGCAAGGGGGTGGGAAGGCCCTCAAATATTGGTGGAAATCTCCTACAGTATTGGA
220 230 240 250 260 270 280

2430 X
GTCAGGAACTAAG
|||||
GTCAGGAACTAAG
290 300

